

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE INFORMÁTICA



## **Solução *Datawarehouse* e *Reporting* Corporativo para o Setor Bancário**

Patrícia Antunes Martins

**Mestrado em Informática**

Trabalho de Projeto orientado por:  
Professor Doutor João Carlos Balsa da Silva  
Engenheiro Rui Manuel Boavida Martinho



*“You don’t need to have all the answers. What you need to do is be curious and open-minded enough  
to learn”.*  
*David Fialkow, Co-founder of General Catalyst*



## Agradecimentos

“Os começos são sempre difíceis. Tão difíceis que muitas pessoas se retraem, reduzindo o inédito e o desconhecido a meras variações sobre o habitual e o conhecido”. É assim, com compreensível orgulho, que pautamos a nossa intervenção nos vários domínios de atuação que nos estão cometidos, nomeadamente na elaboração de trabalhos, exames e, neste caso, de uma tese académica, patrocinando um documento que reúne, num conjunto de textos, o sentir e o olhar de uma pessoa sobre a temática de *Business Intelligence*.

Ao Engenheiro Rui Martinho e ao Professor João Carlos Balsa da Silva, meus orientadores, que coligaram, em tempo recorde, todas as informações requeridas para a elaboração dum trabalho teórico no âmbito do projeto requerido em boa hora concluído e pela opinião sobre a temática abordada, apresento os meus mais sinceros agradecimentos pelo êxito alcançado e, principalmente, pela transmissão de conhecimentos, pela disponibilidade no esclarecimento de dúvidas, pela correção de erros, pelo material bibliográfico fornecido e pelos desafios interessantes propostos ao longo do estágio. Uma palavra especial ao Engenheiro Hugo Figueira e à Engenheira Ana Catarina Miranda, pela colaboração dada nesta singular iniciativa, pelos conselhos, apoio e disponibilidade que revelaram em todas as fases do desenvolvimento deste projeto. Gostaria de deixar um agradecimento especial ao meu colega e parceiro, Pedro Assis, pelo excelente espírito de equipa que manteve, pelo fornecimento de conhecimentos, pelas bases que me permitiram desenvolver um espírito crítico e científico, que em muito ajudaram na realização desta tese e pelo constante apoio e opiniões relativas ao projeto.

Ao Professor António Ferreira, um muito obrigado, pelas bases introdutórias sobre o conceito de *Business Intelligence* e outras temáticas na Unidade Curricular de Integração e Processamento Analítico de Informação.

Ao Duarte Ludovico, um particular obrigado pelo incansável apoio, pela motivação, paciência e, sobretudo, dedicação nas horas mais difíceis. Às minhas amigas, Adriana Alves e Marta Grilo, pelo companheirismo na hora de trabalho e à minha amiga Sofia Gonçalves, pela amizade e disponibilidade de ajuda. À Carolina Barros, Mafalda Barros e Sofia Graça pelas conversas, por me ouvirem e aconselharem e por serem mais irmãs do que amigas.

A todos os colaboradores da Unipartner IT *Services* que me ajudaram direta ou indiretamente, um ilustre agradecimento, pela disponibilidade de espaço e material para a concretização deste projeto e por permitirem a realização de projetos interessantes, produtivos e inovadores. Por último, mas não com menor desempenho, desejo a todos os leitores o mesmo prazer que tive na verdadeira descoberta que para mim representaram estes soberbos e admiráveis textos. A todos, um bem-haja e muito obrigada!



*À minha mãe e irmã.*

vii

“Solução *Datawarehouse* e *Reporting* Corporativo para o Setor Bancário”  
Patrícia Antunes Martins





## Resumo

O projeto insere-se na área de serviços de consultoria da Unipartner IT *Services*, na qual a autora realizou um estágio curricular entre setembro de 2017 e junho de 2018, no âmbito do tema “Solução *Datawarehouse* e *Reporting* Corporativo para o Setor Bancário”.

Este projeto inclui a implementação de uma nova solução de *Business Intelligence* e assenta em quatro objetivos principais: identificação da arquitetura a ser adotada no sistema a implementar, especificação de um modelo de dados para o *Datawarehouse*, desenvolvimento de um processo de *Extraction-Transformation-Loading* (ETL) e execução de relatórios em Microsoft *Power BI*, com base na informação de negócio constante de três fontes de dados operacionais de uma entidade bancária, o Cliente X.

Neste contexto, a autora foi integrada na equipa de projeto, ficando responsável pela análise da informação, definição dos modelos relacionais e produção de relatórios com base numa prova de conceito. A par destas tarefas, desenvolveu-se um processo de ETL, que efetuou grandes transformações aos dados provenientes dos sistemas operacionais, por forma a torná-los mais estruturados e de fácil acesso aos utilizadores finais.

A solução permite ao cliente tomar decisões de negócio mais conscientes, com a devida informação e dados trabalhados, obtendo assim uma visão integrada do negócio.

Este projeto assume, ainda, particular importância no contexto de negócio do Cliente X., uma vez que se trata de uma iniciativa estratégica interna que visa a simplificação e modernização dos instrumentos de apoio à decisão operacional do cliente, em oposição ao sistema manual de criação de relatórios em Microsoft Excel. Ao potenciar o seu valor no contexto do grupo de empresas, do qual faz parte, diminui o tempo de procura de informação e criação de relatórios, aumentando a relação de informação através de métricas. Esta solução foi criada com base em tecnologias Microsoft, nomeadamente *SQL Server* e suas ferramentas.

**Palavras-chave:** *Business Intelligence*; Processo de ETL; *Datawarehouse*; Microsoft *Power BI*; Relatórios em *Power BI*.



# Abstract

This project takes an interest in Unipartner IT Services' Consulting Services, in which the author did an internship from September 2017 until June 2018, connected to the subject “*Solução Datawarehouse e Reporting Corporativo para o Setor Bancário*”.

This paper implements a new Business Intelligence solution and focus on four main purposes: identification of the adopted architecture, definition of a data modelling for Datawarehouse, establishment of an Extraction-Transformation-Loading process and production of reports using Microsoft Power BI, according to the business information of three data sources from a bank sector, Client X.

In this context, the author was incorporated in the project team, taking responsibility for all the information analysis, definition of relational models and development of reports according to a proof of concept. Alongside this, an ETL process was created to transform the data derived from the first dataset, to make them legible and easily accessed by the final users.

This solution allows the client to make the business decision more consciously, with the proper information and worked data, obtaining an integrated vision of the business.

Overall, this project pays a significant attention in Client X's business context, since it is an internal strategy to simplify and upgrade the instruments of decision support system, instead of the manual system of reports designing in Microsoft Excel. By empowering the value of their own business group, allows them to reduce the time of information seeking and reports elaboration, increasing the relationship of information through metrics. To sum up, this solution was based in Microsoft technologies, SQL Server and its tools.

**Keywords:** *Business Intelligence; ETL Process; Datawarehouse; Microsoft Power BI; Reports in Power BI.*



# Índice Geral

Agradecimentos .....	v
Resumo .....	ix
Abstract.....	xi
Índice Geral.....	xiii
Índice de Apêndices.....	xvii
Lista de Figuras.....	xix
Lista de Tabelas .....	xxi
Lista de Código SQL .....	xxiv
Lista de Acrónimos e Abreviaturas .....	xxvi
Capítulo 1 Introdução.....	1
1.1 Descrição e Contexto do Projeto .....	2
1.2 Enquadramento do Projeto .....	3
1.2.1 Enquadramento Académico .....	3
1.2.2 Enquadramento Institucional .....	4
1.3 Motivações.....	6
1.3.1 Motivações do Projeto .....	7
1.4 Finalidade e Objetivos .....	8
1.4.1 Objetivos Específicos do Projeto .....	8
1.5 Planeamento e Execução Inicial do Projeto .....	9
1.5.1 Plano Geral do Projeto .....	9
1.5.2 Plano Detalhado do Projeto .....	11
1.5.2.1 Recursos humanos .....	11
1.5.2.2 Estimação do esforço disponível .....	12
1.5.2.3 Estimação das condições financeiras ou comerciais.....	12
1.6 Principais Contribuições deste Trabalho .....	12
1.7 Notação Adotada .....	13
1.8 Organização e Estrutura do Documento .....	13
1.9 Sumário.....	15

Capítulo 2	Trabalho Relacionado.....	17
2.1	Processamento Analítico de Dados .....	17
2.2	Fontes de Dados Operacionais .....	18
2.3	Descrição do Conceito de <i>Business Intelligence</i> .....	18
2.3.1	<i>Datawarehouse</i> .....	19
2.3.2	Modelação Dimensional .....	21
2.3.3	Processos de ETL.....	23
2.3.4	Microsoft <i>Power BI</i> .....	25
2.3.4.1	<i>Dashboards</i> em <i>Business Intelligence</i> .....	25
2.3.4.2	Relatórios em <i>Business Intelligence</i> .....	26
2.3.4.3	<i>Apps Workspaces</i> e <i>Apps</i> do Microsoft <i>Power BI</i> .....	27
2.3.4.4	<i>Self-Service BI</i> .....	27
2.4	Tipos de Arquitetura: <i>On-prem</i> , <i>Cloud</i> e Híbrido .....	27
2.5	Sumário.....	28
Capítulo 3	Trabalho Realizado.....	31
3.1	Ambiente de Trabalho .....	31
3.1.1	Ferramentas, Linguagens e Tecnologias utilizadas .....	32
3.1.1.1	Microsoft Excel.....	32
3.1.1.2	SQL <i>Server Integration Services</i> .....	32
3.1.1.3	Microsoft <i>Power BI</i> : <i>Power BI Desktop</i> .....	33
3.1.1.4	Linguagem SQL.....	34
3.1.1.5	Linguagem DAX.....	34
3.1.1.6	<i>Master Data Services</i> .....	34
3.1.2	Organização do Trabalho em Equipa.....	35
3.1.3	Processo de Desenvolvimento de <i>Software</i> .....	35
3.2	Identificação de Requisitos.....	37
3.3	Caracterização dos dados do Processo (Modelação Dimensional) .....	38
3.3.1	Fonte de dados A .....	39

3.3.2	Fonte de dados B.....	44
3.4	Arquitetura da Solução e seus principais Componentes.....	48
3.5	Integração .....	52
3.5.1	Processo de ETL .....	52
3.5.1.1	Extração .....	53
3.5.1.2	Transformação .....	55
3.5.1.3	Carregamento.....	57
3.5.2	<i>Datawarehouse</i> .....	58
3.5.3	Avaliação de Segurança dos Dados .....	59
3.6	Camada de Visualização: Relatórios e <i>Dashboards</i> de BI.....	60
3.6.1	Análise Preliminar de Requisitos.....	61
3.6.2	Construção e Implementação dos Relatórios .....	62
3.6.3	Validação do trabalho realizado.....	68
3.7	Plano Efetivo: Análise aos desvios do Planeamento do Projeto .....	68
3.8	Sumário.....	70
Capítulo 4	Conclusão .....	72
4.1	Considerações Finais .....	72
4.2	Principais Contribuições.....	73
4.3	Competências Adquiridas e Apreciação Crítica.....	74
4.4	Principais Dificuldades e Limitações .....	75
4.5	Trabalho Futuro .....	77
	Referências Bibliográficas .....	79
	Apêndices.....	83





# Índice de Apêndices

Apêndice I – Diagrama de Gantt.....	83
Apêndice II – Outras Dimensões existentes na Fonte de Dados A.....	93
Apêndice III – Outras Dimensões existentes na Fonte de Dados B.....	97
Apêndice IV – Campos para a Tabela de Mapeamento para o MDS .....	99



# Lista de Figuras

Figura 1: Digrama do Modelo Organizacional da Unipartner IT Services.....	5
Figura 2: Áreas de Atuação da Unipartner IT Services .....	6
Figura 3: Relação entre os diferentes componentes de um sistema de DW .....	21
Figura 4: Star Schema.....	22
Figura 5: Coordenação entre as diferentes fases do Processo ETL .....	25
Figura 6: Tipos de Arquitetura.....	28
Figura 7: Workflow de Implementação do DW Corporativo .....	31
Figura 8: Processo de desenvolvimento de Software .....	37
Figura 9: Identificação Inicial dos Requisitos (Dimensões, Métricas e Tabelas de Factos) para a fonte de dados A.....	40
Figura 10: Modelo de Dados da Fonte de dados A.....	41
Figura 11: Modelo de Dados da Fonte de dados B .....	45
Figura 12: Arquitetura Física da Solução .....	49
Figura 13: Arquitetura dos <i>Data Flows</i> .....	50
Figura 14: Atualização dos dados no serviço Power BI .....	51
Figura 15: Metodologia do Processo de ETL .....	53
Figura 16: Área de Trabalho do Microsoft Power BI.....	62
Figura 17: Layout definido para os relatórios em Power BI.....	63
Figura 18: Página 1 do Relatório da Fonte de dados A – “Pedidos em Carteira” .....	65
Figura 19: Página 2 do Relatório da Fonte de dados A – “Pedidos Abertos e Fechados” .....	65
Figura 20: Página 1 do Relatório da Fonte de dados B – “Projectos” .....	66
Figura 21: Página 2 do Relatório da Fonte de dados B – “Horas Reportadas” .....	67



# Lista de Tabelas

Tabela 1: Calendarização e Organização do Projeto .....	10
Tabela 2: Tabela de Competências .....	11
Tabela 3: Campos da Dimensão “Cliente” .....	42
Tabela 4: Campos da Dimensão “Estado” .....	42
Tabela 5: Campos da Dimensão “Data” (“Conformed.DimDate”) .....	43
Tabela 6: Campos da Tabela de Factos “Pedidos” .....	43
Tabela 7: Campos da Dimensão “Cliente” .....	45
Tabela 8: Campos da Dimensão “Nível de Serviço” .....	46
Tabela 9: Campos da Dimensão “Tipo de Recurso” .....	46
Tabela 10: Campos da Dimensão “Estrutura do Recurso” .....	46
Tabela 11: Campos da Tabela de Factos “Horas Reportadas” .....	47
Tabela 12: Mapeamento Direto entre os Ficheiros de Histórico e o Modelo de Dados implementado.....	56
Tabela 13: Mapeamento entre os Ficheiros de Histórico e o Modelo de Dados implementado para a Dimensão “Tipo de Pedido”.....	57
Tabela 14: Calendarização da realização dos relatórios preliminar e final .....	70
Tabela 15: Fase 1 do Diagrama de Gantt – Pré-Requisitos .....	84
Tabela 16: Fase 2 do Diagrama de Gantt – Iniciação .....	86
Tabela 17: Fase 3 do Diagrama de Gantt – Desenvolvimento Parte I.....	87
Tabela 18: Fase 3 do Diagrama de Gantt – Desenvolvimento Parte II.....	88
Tabela 19: Fase 4 do Diagrama de Gantt – Operações.....	92
Tabela 20: Fase 5 do Diagrama de Gantt – Lançamento .....	92
Tabela 21: Campos da Dimensão “Stream” .....	93
Tabela 22: Campos da Dimensão “Contabilização” .....	93
Tabela 23: Campos da Dimensão “Plano” .....	93
Tabela 24: Campos da Dimensão “Programa” .....	94
Tabela 25: Campos da Dimensão “Gestor de Projecto” .....	94
Tabela 26: Campos da Dimensão “Número de Projecto” .....	94
Tabela 27: Campos da Dimensão “Motivo” .....	95
Tabela 28: Campos da Dimensão “Gestor da Relação” .....	95

Tabela 29: Campos da Dimensão “Tipo de Pedido” .....	95
Tabela 30: Campos da Dimensão “Dimensionamento” .....	96
Tabela 31: Campos da Dimensão “Recurso” .....	97
Tabela 32: Campos da Dimensão “Projecto” .....	97
Tabela 33: Tabelas de Mapeamento para o MDS .....	99



## Lista de Código SQL

Código 1: Excerto de Código para o Carregamento de Ficheiros de Histórico da Fonte de	
Dados A .....	54





# Lista de Acrónimos e Abreviaturas

## Lista de Acrónimos

## Significado

AD	<i>Active Directory</i>
AOA	<i>Kwanza Angolano</i>
APA	<i>American Psychological Association</i>
APIs	<i>Applications Program Interface</i>
BI	<i>Business Intelligence</i>
BPMN	<i>Business Process Model and Notation</i>
BRL	<i>Real Brasileiro</i>
Cliente X.	<i>Cliente que requisitou o Projeto</i>
CRM	<i>Customer Relationship Management</i>
CSV	<i>Comma-Separated Values</i>
DAT	<i>Data Access Tools</i>
DAX	<i>Data Analysis Expressions</i>
DCL	<i>Data Control Language</i>
DDL	<i>Data Definition Language</i>
DML	<i>Data Manipulation Language</i>
DPA	<i>Data Presentation Area</i>
DQL	<i>Data Query Language</i>
DSA	<i>Data Staging Area</i>
DW	<i>Datawarehouse</i>
EDW	<i>Enterprise Datawarehouse</i>
EIS	<i>Executive Information System</i>
ETL	<i>Extract Transform and Load</i>
FCUL	<i>Faculdade de Ciências da Universidade de Lisboa</i>
GDPR	<i>General Data Protection Regulation</i>
GEPIG	<i>Gabinete de Estudos Pós-Graduados</i>
IaaS	<i>Infrastructure as a Service</i>
ISAM	<i>Indexed Sequential Access Method</i>
MDM	<i>Master Data Management</i>
MDS	<i>Master Data Services</i>
MDX	<i>MultiDimensional eXpressions</i>
ODS	<i>Operational Data Store</i>
OLAP	<i>On-Line Analytical Processing</i>
OLTP	<i>On-Line Transaction Processing</i>
PaaS	<i>Platform as a Service</i>
PMO	<i>Project Management Office</i>
PP	<i>“Pequeno Projecto” (no contexto da fonte de dados B)</i>
PR	<i>“Projecto” (no contexto da fonte de dados B)</i>

RDBMS	<i>Relational Database Management System</i>
RDSMS	<i>Relational Data Stream Management System</i>
SCD	<i>Slowly Changing Dimension</i>
SGBD	Sistema de Base de Dados
SO	Sistemas Operacionais
SP	<i>Stored Procedures</i>
SQL	<i>Structured Query Language</i>
SSAS	<i>SQL Server Analysis Services</i>
SSIS	<i>SQL Server Integration Services</i>
SSMS	<i>SQL Server Management Studio</i>
SSRS	<i>SQL Server Reporting Services</i>
Tis	Tecnologias da Informação
Unipartner IT Services	Empresa onde decorreu o estágio da autora
USD	Dólar Americano
VSAM	<i>Virtual Storage Access Method</i>
XML	<i>eXtensible Markup Language</i>

## **Lista de Abreviaturas**

*vs*

## **Significado**

*Versus*



# Capítulo 1

## Introdução

No mundo empresarial, principalmente no mundo das Tecnologias de Informação (TIs), são constantes as mudanças que ocorrem no ambiente organizacional e diversos os fatores que contribuem para o bom curso de um negócio. Posto isto, é imprescindível que as organizações compreendam as conexões existentes entre todas as vertentes da empresa, por forma a tomarem decisões estratégicas, táticas ou operacionais, facilitando o destaque, crescimento e vantagem competitiva da empresa em relação ao mercado (Vercellis, 2009). Neste contexto, as TIs assumem uma grande importância pois permitem, de forma fácil, rápida e simples, extração, organização, análise e circulação de informação necessária a todos os níveis da empresa. Assim, surge o conceito de *Business Intelligence* (BI), cada vez mais aceite e divulgado na comunidade empresarial, que oferece uma escolha eficaz para que a empresa acompanhe facilmente os seus números de vendas, por exemplo, e possa usá-los como vantagem para um crescimento sustentável, oferecer serviços ou produtos personalizados e estar à frente da concorrência. O sucesso empresarial está relacionado com as estratégias utilizadas que apoiam a tomada de decisão.

Com o decorrer do tempo, as informações começaram a ser geradas e, consequentemente, armazenadas, formando grandes depósitos de dados conhecidos como *Datawarehouse* (DW), e que são, considerados por muitos, a memória da empresa. Assim sendo, o DW (Gardner, 1998) surgiu no mercado com o intuito de organizar os dados de uma organização da melhor maneira, criando uma oportunidade de agilizar e qualificar os processos de decisão e aumentar a sua rentabilidade. Diante deste cenário, a quantidade de informação disponibilizada por estes meios ajuda a otimizar as decisões dos clientes, uma vez que a informação apresentada se encontra organizada e é *user-friendly*, ou seja, adequada, bem estruturada e rápida no sistema de pesquisa. Assim, a verdadeira importância da implementação de um sistema de BI não reside apenas no armazenamento rápido de uma grande quantidade de dados em base de dados, mas também na procura inteligente que permite a flexibilização de consulta e análise de informações de acordo com as necessidades dos utilizadores.

Através da adoção do sistema de BI e do conhecimento dos seus elementos como suporte à gestão estratégica, é possível maximizar a utilidade de informação recolhida e o suporte aos processos de decisão do negócio, obtendo uma visão integrada do mesmo e disponibilizando rapidamente informação relevante para o decisor.

## 1.1 Descrição e Contexto do Projeto

Pretende-se desenvolver um projeto em âmbito empresarial na área da consultoria tecnológica e desenvolvimento que consiste na implementação de uma solução de BI, nomeadamente a criação de um DW corporativo, inserido no estágio curricular realizado na Unipartner IT Services, na unidade curricular do Projeto de Informática. O sistema desenvolvido deverá substituir o sistema manual que está atualmente em funcionamento e aumentar as suas funcionalidades.

O principal objetivo deste projeto remeteu para a exploração dos vários elementos que fazem parte da realidade de BI que, no âmbito do Cliente X., incluem a produção e disponibilização de um conjunto de relatórios dinâmicos, interativos e operacionais por forma a dar resposta às necessidades do Cliente X., de acordo com os requisitos dos seus relatórios mensais disponíveis num documento de prova de conceito, os quais serão utilizados como referência. O intuito deste projeto é assim refletir o estado atual dos relatórios mensais do Cliente X. e, em alguns casos e, dependendo de algumas fontes, ter uma visão integrada do histórico desde o dia 1 de janeiro de 2014.

Para dar resposta efetiva à missão do Cliente X., lançou-se uma iniciativa de modernização na forma de construção de relatórios, que inclui a disponibilização de um *reporting* corporativo, de acordo com a gestão integrada da informação, ao invés do sistema vigente de criação manual de relatórios, no Microsoft Excel, recorrendo a *pivot tables*, de onde se extraem as vistas utilizadas nos diversos relatórios.

O resultado do uso do Microsoft Excel gera, por vezes, informações inconsistentes, imprecisas e lentas, normalmente sob documentos não conexos que possuem, geralmente, algumas diferenças em relação aos originais. Por esta razão, o uso de *pivot tables* por parte de muitas organizações como sistema de *reporting*, não está pronto para analisar grande quantidade de informação, devido às suas características inerentes: não guardam histórico; podem conduzir a erros humanos e gerar perdas para a empresa; são modelados e otimizados para fazer acesso registo a registo; não são capacitados para recuperar e consolidar grandes volumes de dados de forma eficiente; e, de uma maneira geral, criam dificuldades adicionais de integração e análise dos dados, tornando o acesso à informação, uma atividade difícil e, por vezes, impossível. Ainda constituem problemas específicos deste projeto, os seguintes: tempo gasto pelos colaboradores do Cliente X. na produção e consulta de relatórios; desajuste entre os modelos de dados e as necessidades de análise; dispersão de informação nos diversos sistemas de gestão de TIs; tempo excessivo de resposta a análises mais detalhadas / *ad hoc*; ausência de cálculo automático de métricas fundamentais para acompanhamento da atividade do Cliente X.; dificuldade em criar uma visão integrada/consolidada da atividade do Cliente X. devido à restrição no acesso à informação e respetivo tratamento; excesso de recursos na criação dos relatórios.

Esta solução tem como base a informação de negócio constante da agregação de três fontes de dados<sup>1</sup> operacionais pertencentes à atividade desenvolvida pelo Cliente X., no sentido de implementar um DW e incorporar a monitorização e controlo de relatórios e métricas de análise e gestão estratégica numa entidade bancária, através da criação de uma plataforma e na automatização de um *reporting* corporativo em Microsoft Power BI e, por forma a existir uma otimização dos relatórios criados e harmonização dos serviços elaborados, ou seja, para a análise ser feita de uma forma dinâmica e eficaz. A construção de um portal de *reporting* corporativo constitui, assim, um ponto único de acesso, facilita a partilha de informação de gestão nos diversos departamentos do Cliente X. e permite a disponibilização de ferramentas e informação num conceito de *self-service* BI, dotando os utilizadores chave do Cliente X. de capacidades ricas de exploração de informação de uma forma *ad-hoc*.

A solução foi criada com base em tecnologias Microsoft, nomeadamente *Structured Query Language* (SQL) *Server* e suas ferramentas, bem como Microsoft *Power BI*. O uso do *Power BI* possibilita a obtenção de respostas necessárias através da análise do negócio, otimizando decisões, de forma a controlar e a detetar tendências importantes nos dados. Este possibilita, ainda, executar manipulações de grandes conjuntos de dados, permitindo importar mais dados que o Microsoft Excel. Além disso, oferece uma poderosa capacidade analítica, como *Data Analysis Expressions* (DAX) – uma linguagem de expressões que dá suporte à manutenção de dados e à lógica de negócios.

A autora tem a possibilidade de desenvolver o seu projeto e ultrapassar uma série de desafios num ambiente profissional, no seio de uma equipa experiente, jovem, dinâmica, inovadora, com desafios claros e ambiciosos e objetivos planeados, que a irá auxiliar na aquisição dos conceitos necessários para a criação da solução, facilitando a sua integração na vida profissional e proporcionando experiências diversas e enriquecedoras.

A avaliação dos resultados obtidos por parte da equipa de BI, baseou-se essencialmente no produto final resultante dos relatórios criados.

Em suma, a aplicação de um sistema de BI constitui uma alternativa viável por oposição aos manuais, por exemplo, Microsoft Excel. O sistema de BI apresenta alguma maturidade, uma interface com o utilizador aceitável, traduzindo-se numa menor dificuldade de adaptação por parte do mesmo.

Pelo explicado anteriormente, a criação de relatórios interativos e dinâmicos mantém a coerência dos dados e apresenta-os numa forma gráfica de fácil compreensão para os utilizadores, possibilitando a sua exploração de forma autónoma e versátil, com base nas necessidades do Cliente X.

## 1.2 Enquadramento do Projeto

Este projeto tem em conta dois enquadramentos: o académico (Faculdade de Ciências da Universidade de Lisboa) e o institucional (Unipartner *IT Services* e Cliente X.), desenvolvidos de seguida.

### 1.2.1 Enquadramento Académico

O projeto foi desenvolvido no âmbito da unidade curricular – Dissertação/Projeto C (Informática), do 2º ano do plano de estudos do Mestrado em Informática, proposto pela Faculdade de Ciências da Universidade de Lisboa (FCUL), com início em setembro e duração de um ano letivo

---

<sup>1</sup> As três fontes de dados são designadas por A, B e C. Os seus nomes foram omitidos por questões de confidencialidade.

(aproximadamente 9 meses) e com a avaliação de 60 créditos. O trabalho enquadra-se na vertente de Trabalho de Projeto, de acordo com um estágio desenvolvido na Unipartner IT Services. Este estágio teve início no dia 25 de setembro de 2017, com a integração da autora na equipa de projeto a ocorrer apenas no dia 01 de janeiro de 2018. O término do estágio ficou concluído no dia 21 de junho de 2018.

A orientação pedagógica ficou atribuída ao Professor Doutor João Carlos Balsa da Silva, do Departamento de Informática da FCUL. Ao orientador coube mediar a relação entre o aluno e o coorientador da Unipartner IT Services, Engenheiro Rui Martinho, ajudar na decisão do tema, título e plano de trabalho, bem como acompanhar o trabalho desenvolvido e avaliar periodicamente a componente teórica do trabalho de projeto, verificando a qualidade do documento.

## 1.2.2 Enquadramento Institucional

Este projeto insere-se num plano de desenvolvimento gerado pela Unipartner IT Services, líder em consultoria na área das TIs em Portugal, em parceria com uma instituição particular bancária, Cliente X. A inclusão da autora nesta empresa advém de uma relação de parceria na empresa, promovida pela FCUL, responsável pela orientação de um conjunto de trabalhos a realizar por alunos de mestrado num contexto empresarial. Numa fase inicial, a integração da autora no contexto empresarial teve como objetivos conhecer a estrutura da empresa, as competências de cada unidade orgânica, bem como os seus métodos de trabalho, normas internas, política de qualidade e toda a documentação a ser produzida e utilizada no desenvolvimento de projetos. Posteriormente, realizou-se um período de autoformação para compreensão e revisão de conteúdos programáticos lecionados.

A Unipartner IT Services (Unipartner IT Services, S.A, 2018) foi fundada em 2015 por cinco ex-diretores comerciais e de serviços, por meio da reorganização de uma multinacional americana. É uma empresa privada de consultoria e integração de sistemas de informação e que engloba uma equipa multidisciplinar de profissionais, responsável por serviços de TIs no geral, nomeadamente serviços e soluções empresariais inteligentes e económicas no setor público: comercial, indústria e serviços financeiros, de acordo com as necessidades dos clientes, garantindo competitividade, segurança, eficiência de custos e resolvendo questões críticas de negócio. Tem como principais atividades: desenhar, implementar e gerir soluções inovadoras que suportam as médias e grandes empresas e administração pública na sua transformação digital, serviços de aplicações, serviços *Cloud* e de infraestrutura, cibersegurança e gestão de serviços em *outsourcing*.

Esta empresa destaca-se como um dos principais parceiros empresariais da Microsoft, pelo elevado conhecimento e experiência em todas as plataformas e tecnologias Microsoft para o setor empresarial e pelo elevado e diferenciador número de competências que detém. De salientar ainda a larga experiência nacional e internacional da sua equipa de consultoria em projetos nas áreas de grande inovação como: *Social and Customer Engagement*, *Marketing*, *Customer Relationship Management* (CRM) e *Advance Analytics*, *Cloud Platform*, *Data Privacy* e *Cybersecurity*.

O modelo organizacional da empresa (ver Figura 1) inclui:

- *Solutions Area*: os elementos desta área estão, principalmente, focados no *Business Development* de soluções que cruzam várias competências tecnológicas e de soluções verticais. É uma área constituída por um conjunto de *Solutions Leaders*, profissionais altamente qualificados na vertente de tecnologias e sem uma equipa associada.
- *Services Area*: os elementos destas áreas estão principalmente, focados no *Delivery*, *Business Development* focado nas competências e *Sell-On*. Estas áreas de competências



são geridas por *Services Leaders* e *Manager Leaders*, cada um com a responsabilidade de gerir uma equipa focada numa determinada tecnologia Microsoft.

- *Project Management Office (PMO) Area*: Esta área apresenta como função suportar, transversalmente, a gestão de projetos e disponibilizar gestores de projeto para determinados projetos (internos ou externos).

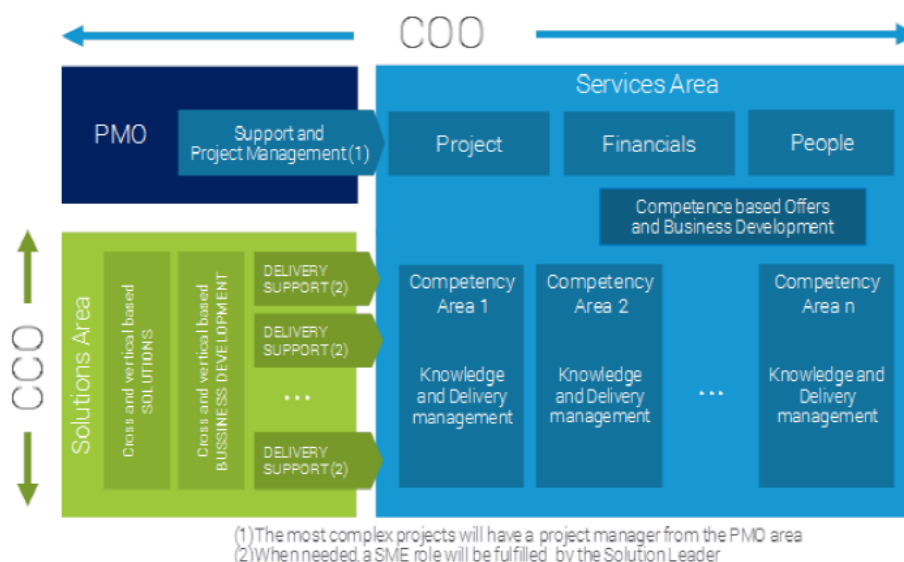


Figura 1: Digrama do Modelo Organizacional da Unipartner IT Services  
Fonte: print screen Unipartner IT Services – Welcome Guide Contractor

Para cumprir o seu objetivo, a Unipartner IT Services integra os seus recursos e investimentos em várias áreas que compreendem o mercado estratégico. O seu foco de trabalho centra-se na comercialização de produtos e serviços, através de venda direta, a nível nacional e internacional, recorrendo a ajuda de distribuidores. Esta empresa trabalha com organizações governamentais, instituições financeiras, serviços financeiros e empresas comerciais como telecomunicações e transportes. As ofertas da Unipartner IT Services incluem consultoria de transformação digital, serviços de aplicações de TI, serviços de *Cloud* e infraestrutura e serviços de gestão e terceirização (Figura 2).

Nos seus projetos tem, atualmente, mais de cento e trinta funcionários e contratados e alguns milhares de parceiros em projetos complementares que oferecem competências e serviços de *outsourcing* (gestão dos *data centers* e dos servidores dos clientes e ambientes computacionais, incluindo processos específicos de negócio, segurança e transformação), integração de sistemas e consultoria (a sua ação decorre em parceria com clientes, permitindo a análise da segurança e eficiência dos custos dos sistemas de TIs até ao momento da intervenção; desenho, integração e modernização das aplicações implementadas, por forma a cumprir os objetivos de negócio), serviços de infraestruturas (desenho e suporte dos serviços de infraestruturas necessários para os serviços de TIs dos clientes, desde redes a dispositivos sem fios), manutenção *core* (manutenção de produtos e sistemas) e tecnologia (desenho e desenvolvimento de servidores e produtos que visam redução de custos e eficiência dos ambientes em *data centers* com o auxílio de gestão de infraestruturas e *Cloud Computing* sempre em prol do utilizador final).



Figura 2: Áreas de Atuação da Unipartner IT Services  
Fonte: print screen Unipartner IT Services – Welcome Guide Contractor

Concluindo, a Unipartner IT Services possui pessoal técnico altamente qualificado e certificado, com uma vasta experiência e competência tecnológicas de elevado nível transacional, focando-se sempre nas carências dos clientes.

Neste contexto, o Cliente X. é uma entidade bancária inserida num grupo de empresas, sob o qual garante eficiência e eficácia na prestação de serviços técnicos e administrativos aos outros gabinetes dos membros, a nível de sistemas de informação e comunicação e outros serviços especializados, nomeadamente no apoio à dinamização do negócio, assessoria fiscal, serviços de suporte à atividade bancária (linha direta) ou através de canais não-presenciais, serviços *On-Line* Particulares e Empresas. Por razões de confidencialidade não será divulgado o nome da organização a que pertence o Cliente X.

### 1.3 Motivações

No desenvolvimento de um projeto, é imperativo a definição de um planeamento, principalmente quando envolve custo, tempo, dinheiro e sacrifício pessoal, lazer e preocupações

diversas. Assim, mais do que uma motivação para a empresa, que visa garantir a satisfação dos seus clientes e a eficácia da implementação do seu produto, também a equipa de trabalho e, especialmente, a autora têm de apresentar motivações e capacidades para a realização do projeto.

A equipa tem como motivações: lidar com novos desafios e expectativas, para atingir um fim comum; conhecer a realidade do negócio do Cliente X.; deter de excelência técnica, experiência e certificações comprovadas, prática de consultoria e seu reconhecimento por parte do mercado; ter capacidade de desenho e implementação da solução aplicacional, permitindo a concretização dos seus objetivos com o maior grau de confiança, qualidade, planeamento e garantia de sucesso; desenhar soluções de BI corporativas, em resposta às necessidades de exploração da informação, análise preditiva e apoio à decisão; fortificar o conhecimento do ecossistema tecnológico do Cliente X.; contribuir para uma maior eficácia da organização do Cliente X.; certificar-se e especializar-se em tecnologia Microsoft *Power BI* e *Self-Service BI*; apresentar uma solução altamente escalável e robusta.

A autora, por sua vez, vê esta iniciativa como uma oportunidade de integrar uma equipa multidisciplinar num contexto empresarial, proporcionando um contato direto com a análise de base de dados e aprofundando os conhecimentos em torno de conceitos emergentes na área de BI.

As motivações do projeto encontram-se detalhadas na secção seguinte.

### 1.3.1 Motivações do Projeto

Como principais motivações para a realização do projeto, foram definidas as seguintes:

- Necessidades recorrentes de *reporting* para suportar a gestão da atividade do Cliente X. (exigência de muitos recursos de vários departamentos para a realização de relatórios mensais);
- Solução para colmatar o tempo excessivo de resposta a análises mais detalhadas \ *ad-hoc* por parte do Microsoft Excel;
- Ausência de cálculo automático de métricas fundamentais para o acompanhamento da atividade do Cliente X.;
- Filtragem de informação desnecessária para a elaboração dos relatórios de apoio à decisão;
- Restrições no acesso à informação e respetivo tratamento;
- Dificuldade em contemplar diferentes perspetivas de análise e informação sobre os vários tipos de atividade potencialmente não uniformizada;
- Desajuste entre os modelos de dados e as necessidades de análise;
- Dificuldade em criar uma visão integrada/consolidada da atividade do Cliente X. (os dados necessários encontram-se acomodados a várias fontes de dados, sendo que a relação de informação, muitas vezes, se torna mais difícil e demorada).

## 1.4 Finalidade e Objetivos

Um dos componentes essenciais de um projeto de conclusão de mestrado são os objetivos gerais e específicos, responsáveis por direcionar a pesquisa e obter os resultados esperados com o trabalho académico. Do ponto de vista da empresa, os seus principais objetivos assentam: na oferta de soluções robustas que acrescentem valor às operações dos clientes, melhorando a eficiência, eficácia e efetividade das análises de negócio em que o Cliente X. participa; na harmonização com a arquitetura do Cliente X. e compatibilidade futura com o ecossistema tecnológico; na criação de uma solução flexível e configurável para acomodar novos requisitos de exploração da informação numa perspetiva de *self-service*. Da perspetiva da equipa, os objetivos propostos centram-se: no conhecimento das variantes dos negócios, aplicações, infraestruturas e modelos de dados para suporte e evolução do método de produção de relatórios, como base para a definição da construção do *reporting* corporativo que melhor se enquadra ao Cliente X. (diagnóstico do problema); na criação de uma solução de apoio à decisão utilizando diversas tecnologias Microsoft na área de BI, para a integração e modelação de dados e apresentação da informação através de relatórios; na capacidade para assegurar a manutenção futura da solução. Para a autora, os objetivos a que se auto propôs são os seguintes: aprofundar conhecimentos adquiridos durante o mestrado; formação *on job* nas soluções utilizadas no projeto; experiência de trabalho em equipa e em cliente; introduzir na sociedade informática, os conhecimentos e resultados obtidos neste trabalho de projeto.

De uma forma geral, **o objetivo principal do projeto** consiste no desenvolvimento e implementação de um DW e *reporting* corporativo num cliente do setor bancário em ambiente real (Cliente X.), que agrega as diferentes fontes de dados e os relatórios em Microsoft Excel existentes, usando as capacidades de exploração e manipulação de dados inerentes a ferramentas de BI. Este objetivo assenta na construção de relatórios personalizados e métricas base de gestão do Cliente X., os quais garantem, de forma automática e controlada, resposta às áreas de negócio beneficiárias do Cliente X. Com este sistema avançado de análise de informação, o utilizador final terá autonomia total para construir os seus gráficos, relatórios e análises de gestão personalizadas. Num contexto mais centrado, o objetivo geral do projeto direciona o trabalho através dos objetivos específicos clarificados imediatamente.

### 1.4.1 Objetivos Específicos do Projeto

Os objetivos do projeto estão numerados pois constituem a base de trabalho para este trabalho de projeto, sendo que a cada tarefa do planeamento será correspondido um determinado objetivo:

1. Levantamento de requisitos e Modelação das fontes de dados, assentes em análises dos dados nos ficheiros, em Microsoft Excel que permitam responder a todas as análises que são feitas nos relatórios mensais, antes de se partir para uma implementação definitiva;
2. Desenvolvimento de um processo de *Extract Transform and Load* (ETL) para as várias fontes de dados e em conformidade com um modelo de dados previamente definido e avaliado pelo cliente, que permita a extração e transformação dos diferentes dados;
3. Operações de tratamento e filtragem de dados guardados nos vários sistemas de informação existentes no Cliente X., com o objetivo de visualização gráfica dos dados e tendências do

negócio, por forma a simplificar a tomada de decisões, garantir eficiência nos tempos de resposta e que reflitam o estado atual e histórico das fontes de dados;

4. Definir um repositório de dados históricos (DW) para as três fontes de dados e, para auxiliar os colaboradores, contendo informação relevante para a tomada de decisões;
5. Consolidar e reunir a informação referente às três fontes de dados (DW), de modo a que os dados necessários não se encontrem espalhados pelas várias fontes por forma a ser mais fácil a realização dos relatórios de apoio à decisão, levando à consulta e filtragem de informação a partir de outros sistemas de informação internos;
6. Produção de relatórios automáticos de apoio à decisão e de painéis interativos com dados analíticos a serem utilizados pelos diversos departamentos do Cliente X. recorrendo à ferramenta Microsoft *Power BI*, com informação detalhada de dados de modo a fornecer ao cliente representações gráficas da informação e para permitir a exploração dos dados de forma autónoma e versátil com base nas necessidades da direção do Cliente X.;
7. Promover o conceito de *self-service BI*, dotando os utilizadores do Cliente X. de capacidades ricas de exploração de informação de uma forma *ad hoc*, incluindo a criação de relatórios com novas perspetivas da atividade do Cliente X.

## 1.5 Planeamento e Execução Inicial do Projeto

As atividades de um projeto precisam de ser definidas para que o projeto se torne mais realista e factível. Desta forma, esta etapa é uma das mais importantes no ciclo de vida de um projeto e torna-se necessário criar um calendário com datas de início, fim, durações e descrição de atividades de modo a contemplar os objetivos, âmbito do projeto, tempo estimado, viabilidade, metas, restrições e recursos necessários a considerar no seu desenvolvimento.

### 1.5.1 Plano Geral do Projeto

O projeto de implementação de um DW corporativo enquadra-se numa estratégia de construção global de BI a desenvolver de forma faseada, potenciando a obtenção progressiva e alargada de informação crítica à gestão operacional e estratégica do Cliente X.

Numa primeira fase do plano geral inicial, ocorre a integração da autora na organização, apresentação do projeto no qual a autora foi inserida, preparação inicial, estudo da arquitetura existente e formação inicial. Juntamente com o gestor do projeto, foram identificadas as funcionalidades a desenvolver pela autora e realizada a respetiva análise funcional. Este projeto foi, constantemente, submetido a testes, correções e atualização da documentação técnica.

De uma forma geral, o plano geral e inicial de trabalhos para o projeto envolveu as seguintes etapas e sub-etapas com as respetivas datas de início e fim e duração correspondente (Tabela 1). O plano de trabalhos deste projeto iniciou-se em novembro de 2017 com a análise, tratamento de dados e elaboração de relatórios da fonte de dados C, razão pela qual o plano de trabalhos tem início na semana 9.

Tabela 1: Calendarização e Organização do Projeto

Início	Fim	Atividade
<b>1-jan-18</b>	<b>5-jan-18</b>	<b>Semana 9</b>
		<i>Kick-Off</i> do Projeto BI – Informação de Gestão.
		Análise Funcional <i>High-Level</i> (Recolha de dados e requisitos funcionais).
<b>8-jan-18</b>	<b>19-jan-18</b>	<b>Semana 10 e 11</b>
		Construção da 1ª <i>Release</i> do Relatório Mensal.
		Modelo <i>SandBox</i> das 3 fontes de dados + Relatórios.
		Elaboração do Modelo Relacional.
<b>22-jan-18</b>	<b>26-jan-18</b>	<b>Semana 12</b>
		Análise Funcional Detalhada.
		<i>Workshop</i> revisão requisitos Fonte de dados A + Segurança.
		<i>Workshop</i> revisão requisitos Fonte de dados B + Segurança.
		<i>Workshop</i> revisão requisitos Fonte de dados C + Segurança.
		Revisão do documento técnico.
<b>29-jan-18</b>	<b>9-fev-18</b>	<b>Semana 13 e 14</b>
		Camada de Ingestão de Dados.
		Construção da <i>Staging</i> , DW.
		Processo de ETL de dados: Fonte de dados A.
		Processo de ETL de dados: Fonte de dados B.
		Processo de ETL de dados: Fonte de dados C.
<b>12-fev-18</b>	<b>16-fev-18</b>	<b>Semana 15</b>
		Camada de Processamento de Dados.
		Desenho da Aplicação.
<b>19-fev-18</b>	<b>2-mar-18</b>	<b>Semana 16 e 17</b>
		Camada de Repositórios de Dados.
		Modelo analítico atividade planeada – Fonte de dados A + Fonte de dados B.
		Modelo analítico atividade corrente – Fonte de dados C.
<b>5-mar-18</b>	<b>16-mar-18</b>	<b>Semana 18 e 19</b>
		Camada de Apresentação – Portal de <i>Reporting</i> ( <i>Power BI</i> ).
		<i>Power BI Governance</i> .
		Criação e Disponibilização dos relatórios finais da fonte de dados A.
		Criação e Disponibilização dos relatórios finais da fonte de dados B.
		Criação e Disponibilização dos relatórios finais da fonte de dados C.
		<i>Dashboards</i> consolidados, por parte do departamento de Planeamento e Controlo do Cliente X.
<b>19-mar-18</b>	<b>23-mar-18</b>	<b>Semana 20</b>
		Disponibilização no ambiente de qualidade.
		Procedimento de instalação incluindo <i>scripts</i> de Instalação.
		Eventuais ajustes.

26-mar-18	30-mar-18	Semana 21
		Disponibilização no ambiente de produção.
		Entrada em Produção.
		Apoio Pós-Produção.

Fonte: Elaborado pela autora

## 1.5.2 Plano Detalhado do Projeto

O plano detalhado do projeto inclui a decomposição de tarefas e a sua breve descrição, de acordo com o esquematizado no diagrama Gantt (Keunecke, 2016), que se encontra em apêndice (Apêndice I – Diagrama de Gantt) e compreende 5 fases essenciais: Pré-requisitos (Fase 1); Iniciação (Fase 2); Desenvolvimento parte I e II (Fase 3); Operações (Fase 4); Lançamento (Fase 5). Numa fase inicial, ocorreu uma fase de autoformação, tendo incidido principalmente na aprendizagem das ferramentas a utilizar no desenvolvimento do projeto e outras tecnologias periféricas a serem utilizadas.

Todas as atividades que referenciam pontos de situação internos ou externos (dos quais o Cliente X. é participante) servem para monitorizar e controlar o bom curso do projeto.

### 1.5.2.1 Recursos humanos

A equipa de projeto é constituída pela equipa técnica da Unipartner IT *Services* que envolve um diretor de projeto, um gestor de projeto e dois consultores de BI. Da parte do Cliente X., foram designados interlocutores chave, normalmente um por unidade orgânica para as três fontes de dados, perfazendo um total de três e um diretor de projeto, gestor de projeto, engenheiro da solução, equipa de infraestruturas e equipa de testes. O modelo de gestão, coordenação e controlo do projeto foi monitorizado em concordância com as normas de trabalho definidas pela Unipartner IT *Services*. O pessoal técnico afeto à operação engloba elementos internos do Cliente X. e aquisição de serviços externos da Unipartner IT *Services*.

A tabela de competências pertencente à autora com os recursos de *software* para o desenvolvimento do projeto encontra-se disponível na Tabela 2.

Tabela 2: Tabela de Competências

Ferramentas, Linguagens e Tecnologias	Domínios
Microsoft Power BI ( <i>App, Workspace, Power BI Desktop</i> )	Razoável
Linguagem SQL	Razoável
Linguagem DAX	Nenhum
SQL Server Integration Services (SSIS)	Razoável
Master Data Services (MDS)	Nenhum
Microsoft Excel	Bom

Fonte: Elaborado pela autora

### 1.5.2.2 Estimação do esforço disponível

A abordagem de calendarização apresentada dá o projeto por terminado num total de 21 semanas (sensivelmente 5 meses), aproximadamente, 840 horas de trabalho (8 horas de trabalho diários x 5 dias úteis x 21 semanas), com início previsto a 01 de novembro de 2017 e fim a 30 de março de 2018, sendo que a autora integrou oficialmente o projeto no dia 8 de janeiro de 2018. O local de desenvolvimento do projeto foi no Cliente X., com deslocações esporádicas à Unipartner IT *Services* para reuniões mensais de projeto e pontos de situação internos.

### 1.5.2.3 Estimação das condições financeiras ou comerciais

Por motivos de segurança e em defesa do princípio da confidencialidade (Portugal, Constituição da República Portuguesa, 2013) nomeadamente na proteção dos dados pessoais, do sigilo fiscal e na proteção dos direitos de propriedade industrial, pelo Regulamento da Política de Valorização de Conhecimento da Faculdade de Ciências da Universidade de Lisboa, Despacho n.º 2467/2017, de 22 de março, Diário da República, 2.ª Serie, N.º 58 (Portugal, Constituição da República Portuguesa, 2017), que é do conhecimento da Unipartner IT *Services* e das restantes partes, serão omitidos nomes de clientes, nomes de elementos de equipa, nomes de responsáveis de entidade que compõem o Cliente X., informações técnicas, comerciais, económicas, financeiras e administrativas quer da Unipartner IT *Services* quer do Cliente X.

## 1.6 Principais Contribuições deste Trabalho

Neste ponto do projeto, são inseridas as contribuições de cada uma das partes do ponto de vista da informação e dos recursos necessários para a realização do projeto.

Inicialmente, foi discutida a proposta do projeto com o responsável do projeto da Unipartner IT *Services* que apresentou ao Cliente X., a definição do projeto, os objetivos, o âmbito, a metodologia que se proponha utilizar e a estratégia a adotar, assim como as várias fases do trabalho, estimativa de duração do mesmo e a equipa a envolver na execução do projeto.

O projeto foi desenvolvido por uma equipa composta por duas pessoas, em que os dois participaram na totalidade da definição da solução. A meio do desenvolvimento da solução, surgiu uma terceira pessoa responsável pela definição dos relatórios em Microsoft *Power BI*. Em relação à implementação da solução, as tarefas foram divididas pelos dois elementos. Aquando da entrada da autora no projeto, no dia 8 de janeiro de 2018, já tinha sido concluída a primeira abordagem à fonte de dados C (contém informação sobre pedidos de serviço e incidentes abertos no contexto da atividade corrente do Cliente X.), cuja análise de requisitos ocorreu no dia 01 de novembro de 2017. Assim sendo, a autora ficou encarregue do desenvolvimento dos processos de ETL, construção do DW e relatórios em Microsoft *Power BI* para a fonte de dados A (contém informações sobre pedidos que suportam a atividade planeada do Cliente X.), enquanto o parceiro de equipa, tratou da fonte de dados B (contém a informação sobre o planeamento de projetos, quando aplicável, e o registo dos recursos que lhes estão alocados).

Neste projeto, a autora participou nas várias fases de definição da solução, as quais constituem os principais objetivos do projeto, referenciados anteriormente, e da sua implementação: análise funcional das fontes de dados operacionais, principalmente da fonte de dados A e levantamento dos respetivos requisitos; apresentação de abordagens/soluções inovadores para os desafios identificados; definição dos modelos de análise dos dados pretendidos (desenho da solução) para a fonte de dados A e



B; definição e desenho da arquitetura de solução, com base nas regras de segurança do Cliente X.; implementação dos processos de ETL para as fontes de dados A e B; acompanhamento da avaliação de desempenho dos processos construídos e identificação de falhas; elaboração de documentação relativamente à solução implementada (documentos funcionais das fontes de dados e documento técnico); construção do DW referente aos casos de uso existentes no documento da prova de conceito: construção da camada de processamento, ingestão e repositório de dados; criação de uma plataforma de *reporting* corporativo, através da criação de relatórios e de métricas de negócio exigidas pelo Cliente X. em Microsoft *Power BI*; disponibilização da solução nos ambientes de qualidade e produção através de *scripts* definidos pelo Cliente X. e apoio pós-produção (colaboração em fases de melhoria e otimização da solução implementada).

## 1.7 Notação Adotada

Para a nomenclatura do presente documento, ficou definido como língua, o português, de Portugal, que contempla as alterações propostas pelo Novo Acordo Ortográfico. A apresentação dos estrangeirismos, por exemplo, Inglês, encontra-se a itálico, tal como o termo *Business Intelligence*. A **negrito** ou a sublinhado, estão os termos, palavras ou expressões que se pretendem destacar, sendo que a **negrito**, estão as mais importantes e a sublinhado, as de menor importância. Note-se que todos os termos e conceitos próprios do projeto se encontram em português de Portugal, mas não contemplam as alterações propostas pelo Novo Acordo Ortográfico, por ser a política de nomenclatura adotada pelo Cliente X. Ainda, todos os campos referentes a dimensões, tabelas de Factos e métricas são colocados entre aspas, como por exemplo: da fonte de dados A, são definidas as dimensões “Cliente” e “Estado”.

A estrutura adotada para a apresentação desta dissertação respeita o guia de dissertação disponibilizado pela coordenação do Mestrado em Informática.

Por questões de confidencialidade com a Unipartner IT Services e o Cliente X., os nomes referentes a clientes, elementos da equipa, informações técnicas, comerciais, económicas, financeiras e administrativas serão substituídas por siglas, por exemplo, Cliente X.

As referências bibliográficas seguiram as normas da *American Psychological Association* (APA), 6ª edição (Silva, 2014).

## 1.8 Organização e Estrutura do Documento

Para atingir os objetivos e de acordo com a metodologia seguida, o presente trabalho de projeto encontra-se organizado em quatro capítulos, subdivididos em diversos tópicos, com a seguinte estrutura:

- Capítulo 1 – Introdução:

Este capítulo compreende o enquadramento conceitual, onde são descritos todos os conceitos inerentes ao estudo (introdução ao conceito de BI e DW, descrição do projeto, identificação do problema e delimitação do âmbito), uma proposta de resolução para o problema em estudo e principais limitações do mesmo, enquadramento académico e institucional do projeto, motivações da empresa, do projeto, da equipa e da autora, realçando a justificação e pertinência da temática para o Cliente X. e os objetivos da empresa, do projeto, da equipa e da autora. Ainda neste capítulo, é descrito o planeamento geral e

detalhado inicialmente proposto para a execução do projeto, as contribuições alcançadas e a notação adotada na escrita do presente documento, terminando com a organização e estrutura do documento.

- Capítulo 2 – Trabalho Relacionado:

Este capítulo diz respeito ao enquadramento metodológico de todos os trabalhos relacionados, abordando o estado de arte, tendo em conta os trabalhos já desenvolvidos na área, que suportam e incentivam o desenvolvimento do projeto e o esclarecimento de conceitos importantes mencionados no projeto no âmbito de BI: descrição do processo de BI, referência aos benefícios, objetivos de utilização de BI e conceitos associados como o *self-service* BI, desvantagens, vantagens e fracassos, processos de ETL, DW, *dashboards* e relatórios de BI e tipos de arquitetura existentes, por exemplo. O capítulo é necessário para a compreensão do projeto e da sua relação com a construção e implementação de um sistema de base de dados.

- Capítulo 3 – Trabalho Realizado:

Neste capítulo, figura a apresentação e desenvolvimento do problema e de todo o trabalho realizado durante o estágio na Unipartner IT Services. Ainda neste capítulo, são descritos os casos de uso e a solução adotada face às exigências do Cliente X., bem como uma explicação sucinta e descrição das principais características de todas as ferramentas, tecnologias, linguagens, ambientes de desenvolvimento utilizados no projeto, processo de desenvolvimento de *software* e visão geral da sua arquitetura. Também se inclui neste capítulo, todas as funcionalidade e passos de implementação de um sistema de BI, descrevendo em detalhe, a construção do DW para o Cliente X., passando pelos passos de extração, transformação e carregamento, característicos do processo de ETL e produção dos relatórios dinâmicos e *dashboards* em Microsoft Power BI. O capítulo termina com algumas considerações relevantes e explicação dos desvios ao plano inicialmente proposto para o projeto.

- Capítulo 4 – Conclusões:

Este capítulo discute a importância e relevância do projeto e ressalva as principais contribuições e conclusões alcançadas a nível prático e teórico sobre o trabalho realizado, expondo as limitações sentidas na sua concretização, terminando, com sugestões para trabalho futuro, decorrente do projeto elaborado e evidenciando apontamentos interessantes que se recomendam para novos projetos.

Ainda na estrutura do documento, surgem dois pontos importantes:

- Referências Bibliográficas: são essenciais para a compreensão teórica e prática do projeto, por serem a fonte das citações colhidas;
- Apêndices: são criados pela autora e contêm informação menos relevante do que o corpo do texto principal, servindo de suporte ao projeto:
  - O Apêndice I refere-se ao diagrama de Gantt que organiza por fases e subfases o plano detalhado do projeto.
  - O Apêndice II diz respeito às dimensões que não possuem hierarquia pertencentes à fonte de dados A.
  - O Apêndice III diz respeito às dimensões que não possuem hierarquia pertencentes à fonte de dados B.
  - O Apêndice IV diz respeito aos campos das três fontes de dados que precisavam de ser decodificados por tabelas de mapeamento no *Master Data Services* (MDS).

## 1.9 Sumário

Neste capítulo, é descrito o enquadramento conceptual do estágio efetuado, pela autora, na equipa de BI, seguindo-se, depois, noutros capítulos, o enquadramento teórico, relato das atividades desenvolvidas, e conclusões retiradas do estágio.

Este projeto pautou-se por uma interpretação de fenómenos com base na análise de dados, passo fundamental para maximizar a informação disponível e apoiar o processo de tomada de decisão, procurando compreendê-los a partir das necessidades, frustrações, disparidades da singularidade e, simultaneamente, complexidade dos dados analíticos do Cliente X. Assim, mais do que a criação de uma estratégia de armazenamento de dados, foi necessário a definição de um modo de consulta inteligente e procura refinada para que o processo de decisão ocorresse com menor probabilidade de erros.

Em cada uma das componentes representadas neste ponto, apresentaram-se as noções fundamentais a observar no problema em questão e sugeriram-se um conjunto de motivações e objetivos a responder, com vista a aferir a qualidade do trabalho produzido. Ainda neste capítulo, foi descrito o enquadramento institucional e académico do projeto, o plano de trabalho e sua execução, as principais contribuições do projeto e a notação utilizada na produção do documento e do projeto. Concluiu-se o capítulo com a apresentação da organização e estrutura do documento.



# Capítulo 2

## Trabalho Relacionado

O trabalho relacionado é uma das partes mais importantes de um trabalho de projeto, uma vez que faz referência ao que já foi descoberto sobre o assunto abordado ou até mesmo o que falta descobrir, filtrando as investigações e paradigmas mais adequados para a correta compreensão do projeto e entendimento dos capítulos seguintes.

### 2.1 Processamento Analítico de Dados

Os sistemas de processamento analítico são uma tecnologia de BI utilizada para explorar um DW e, para apoiar as organizações nas análises efetuadas. Os sistemas *On-Line Analytical Processing* (OLAP) (Haiyan, Zhenyuan, 2010) permitem recuperar e analisar informação a partir dos dados, geralmente num DW ou *Data Marts* através da criação dos cubos multidimensionais. Enquanto um DW é utilizado para armazenar a informação, os sistemas OLAP são utilizados para recuperá-la. Estes sistemas apresentam os dados utilizando Factos, dimensões, cubos e hierarquias de conceitos através de diversas técnicas e operações de visualização.

Os OLAP (Chaudhuri, Dayal, Narasayya, 2011) são sistemas dinâmicos e multidimensionais de apoio à decisão que reúnem dados provenientes de várias fontes, facilitam a navegação e análise dos dados e correspondem a processos longos e exploratórios de leitura de dados. A informação analisada pela ferramenta OLAP é proveniente das bases de dados multidimensionais (DW), o que possibilita aos utilizadores, a combinação da informação de muitas maneiras diferentes.

No fundo, estes sistemas geram, manipulam e armazenam dados. Como principais características, temos as seguintes: orientados a análises; envolvem processamento exploratório; acesso a muitos registos de cada vez; exigem nomes inteligíveis; constituem dados detalhados e agregados; é possível ter acesso à evolução histórica; necessitam de atualizações planeadas e periódicas; provêm de fontes de dados internas e externas; exibem relatórios personalizados; detêm de uma disponibilidade relaxada; apresentam dados históricos; estrutura otimizada para atualizações.

As ferramentas OLAP permitem fazer várias ações sobre os cubos de dados e enquadram-se em modelos simples de análise e identificação de padrões nos dados. A organização dos dados, imposta pelos esquemas multidimensionais, permite que estes sejam analisados sob diferentes perspetivas. Contudo, as análises que recorrem a este tipo de sistemas são baseadas em hierarquias de conceitos para consolidar os dados e para criar visualizações ao longo das dimensões de um DW (Vercellis, 2009). Para além destas funcionalidades, as ferramentas de OLAP permitem normalmente visualizar gráficos ilustrativos dos dados representados nas dimensões do cubo e visualizar conjunto de dados organizados em cubos multidimensionais. A sua finalidade é a de reunir conjuntos de dados de forma organizada e hierárquica fornecendo uma melhor organização dos dados e ajudando a uma melhor compreensão por parte dos utilizadores.

Os sistemas OLAP lidam bem com modelos de dados simples e desnormalizados, em que existem poucas tabelas de grande dimensão, reunindo dados de múltiplas fontes (*star schemas*). No cômputo geral, estes sistemas são desenhados para interrogações exploratórias e analíticas ao alcance

dos decisores, sem seguirem padrões (só é possível por a carga de sistema ser denominada por leituras) e com menos junções de tabelas, menos nomes e mais inteligíveis. O facto destes sistemas não seguirem padrões implicam muita redundância, muitos índices e muitos dados agregados pré-calculados.

## 2.2 Fontes de Dados Operacionais

*On-line Transaction Processing* (OLTP) (Kimball, Ross, 2013), ao contrário dos sistemas OLAP, são fontes de dados que armazenam sistemas de processamento de transações que fazem os dados entrar na organização, garantem o seu funcionamento e correspondem a processos curtos e repetitivos de escrita e leitura de dados (inúmeros *selects*, *inserts*, *updates* e *deletes*). Como principais características dos Sistemas Operacionais (SO), temos as seguintes: orientados a transações; envolvem processamento repetitivo; permitem que poucos dados sejam acedidos, sem redundância; exigem nomes curtos e códigos; constituem dados detalhados; os dados acedidos contemplam o estado atual; os dados são atualizados em contínuo; provêm de fontes de dados internas; exibem relatórios pré-definidos; e detêm de uma alta disponibilidade; baseados em assuntos ou negócios; pequena quantidade de dados é usada num processo; baixa ou pequena probabilidade de acesso; não permitem registo histórico, normalmente limitado a dados recentes ou atuais; formato dos dados é, maioritariamente, uniforme nas diferentes fontes de dados. Em sistemas OLTP, os dados são extraídos das fontes por intermédio de processos de ETL e são associados a uma memória de curto prazo, uma vez que guardam apenas a informação recente em comparação com um DW que guarda informação histórica.

Os sistemas OLTP lidam bem com modelos de dados complexos e normalizados, em que existem muitas tabelas, pequenas, poucas colunas (*snowflakes*) evitando anomalias na escrita de dados. Durante a carga de sistema típica, estes sistemas estão afinados para responder rapidamente. Fora da carga de sistema típica, apresentam interrogações analíticas custosas, muito lentas e difíceis de formular pelos decisores, pois exigem muitas junções de tabelas.

## 2.3 Descrição do Conceito de *Business Intelligence*

Antes apelidado (na década de 80) de *Executive Information System* (EIS), o sistema informático (Watson, Houdeshel, Rainer 1997) que simplificava o acesso a informação interna e externa relevante para os fatores críticos de sucesso dos executivos, apresentava como principais características, as seguintes: adaptáveis a necessidades individuais; análise e navegação exploratória dos dados; aplicação de filtros e agregações hierárquicas; cruzamento de várias fontes de dados; acompanhamento de tendências e sinalização de exceções; apresentação gráfica, tabular e textual da informação; tempo de resposta reduzido e/ou bem conhecido; pouco ou nenhum treino para serem usados. Ainda, com a adoção de EIS a qualidade de gestão estratégica melhorou nas organizações, por intermédio da implementação de novas tecnologias e técnicas para a extração, transformação, processamento e apresentação de dados no processo de tomada de decisão.

Face a todas estas características e na melhoria destas, surge o conceito de BI em 1989, por Howard Dresner (BI LIVRE, 2017), definido como o conjunto de “conceitos e métodos que melhoram a tomada de decisões de negócio, usando sistemas de suporte baseados em factos, ou seja, sistemas de suporte à decisão”. Deste modo, o conceito de BI não tem a finalidade de decidir, apenas reúne os elementos visuais capazes de suportar a tomada de decisões. Atualmente, está a usufruir de um maior reconhecimento por parte das empresas e a sua implementação é cada vez mais impactante e o

reconhecimento das suas potencialidades mais notório. A entrada do conceito no contexto empresarial acontece pelo relevo ganho pelos sistemas de informação na atividade das empresas, passando a ser essencial, a sua utilização, nas atividades laborais.

Os sistemas de informação permitem auxiliar a atividade das empresas, através do registo de dados. Os dados registados e guardados num sistema de gestão de base de dados permitem gerar informação, e é esta a informação que permite após ser trabalhada, gerar métricas para a otimização do negócio. Foi com base na necessidade de implementar processos de análise de informação e a partir dela extrair conhecimento que nasceu o termo BI. O conceito de BI é mais vantajoso quanto maior for a quantidade de dados relacionados com a atividade empresarial, disponíveis.

De um modo geral, um sistema de BI serve para (Ramos, Santos 2009): analisar dados passados ou atuais; prever fenómenos e tendências; analisar e comparar dados do passado com novos dados de forma a perceber o que mudou; permitir o acesso *ad hoc* a dados para responder a questões que não se encontram pré-definidas; analisar a organização de modo a obter um conhecimento mais profundo das suas atividades. A inteligência para o negócio (BI) é uma mistura de tecnologias, nomeadamente DW, OLAP e ferramentas de utilização. Assim sendo, os dados são transferidos dos sistemas originais para um local onde poderão ser analisados, o DW, permitindo os acessos a todos os dados de uma empresa. A implementação de um sistema deste tipo pode contribuir de forma eficiente e precisa no desenvolvimento de uma organização, criando o conhecimento necessário para projetos futuros, de modo a minimizar falhas do passado e maximizar o seu desempenho.

### 2.3.1 *Datawarehouse*

O sucesso alcançado pelo BI traduz-se, fulcralmente, na consolidação de dados de diferentes fontes oriundas dos sistemas operacionais da empresa para o DW (Ranjan, 2009), ou seja, constitui o ponto central da arquitetura de processamento de informações para sistemas informáticos modernos suportando o processamento de informação de um sistema de apoio à decisão através da integração de dados corporativos e históricos. Nesta linha de pensamento, a vantagem competitiva alcançada pelo BI acontece através da análise de dados consolidados no DW que é um elemento chave na arquitetura de uma ferramenta de BI. O DW (Kimball, Ross, 2013) é a implementação física de apoio a decisões, uma área de apresentação de dados integrados, temáticos, não voláteis, dependentes de tempo, que disponibiliza aos executivos, as informações necessárias à tomada de decisões estratégicas. Desta forma, possibilita análises exploratórias e estudos de tendências, pelo que terá de ser viável e exequível e deverá aumentar a rentabilidade dos negócios.

Os principais objetivos dos DW (Kimball, Ross, 2013) são: armazenamento de grande quantidade de dados de grandes dimensões que guardam dados de vários processos de negócio; simplificação e diminuição do tempo de acesso exploratório aos dados; consolidação de dados coerentes de várias fontes num repositório; adaptação à mudança, por possibilitar a manutenção de um histórico e permitir alterações graduais sem comprometer a informação; respostas rápidas a interrogações complexas, através do armazenamento de dados agregados pré-calculados, descartando as normalizações que a arquitetura de bases de dados relacionais exige; aumento do desempenho das consultas às bases de dados.

Os elementos mais importantes de um DW (Kimball, Ross, 2013) são 3: *Data Staging Area* (DSA) que guarda os dados extraídos dos sistemas operacionais, a *Data Presentation Area* (DPA) onde os dados prontos são carregados e a *Data Access Tools* (DAT) que apoiam a tomada de decisões

analíticas. As especificidades da construção de um DW contêm o pré-processamento e transformação de dados, análise estatística, construção de diagramas de interligação de fontes de dados e modelação dimensional dos dados analisados.

Uma das maiores preocupações em sistemas de DW é como planejar a sua construção. Duas abordagens são normalmente aceites na implementação destes sistemas, uma do tipo *top-down* e outra do tipo *bottom-up*. A primeira abordagem, sugerida por Inmon (2005), subdivide-se em duas etapas. A primeira etapa consiste na definição do esquema global do DW e a segunda baseia-se na implementação de *Data Marts* de acordo com as necessidades e características das várias unidades de negócio da organização. Esta implica mais tempo de desenvolvimento e maior risco de não ser concluída dentro do prazo estabelecido (Vercellis, 2009). A segunda abordagem, sugerida por Kimball e Ross (2013), baseia-se na experimentação e desenvolvimento de protótipos (Ramos, Santos, 2009). É uma abordagem mais simples e incremental. O objetivo passa por construir esquemas, individuais, de cada *Data Mart*, tendo em consideração as necessidades de cada unidade de negócio. Estes esquemas devem ser modelados tendo em vista a futura unificação dos mesmos, de modo a obter-se um DW organizacional. Geralmente esta abordagem é mais rápida, fornece resultados mais tangíveis, mas carece claramente de uma visão conjunta de todo o sistema a ser desenvolvido (Vercellis, 2009). O tipo de abordagem depende muito das necessidades organizacionais.

De um modo geral, o DW centraliza dados provenientes de diversas fontes de uma organização (*data source*) e utiliza processos consistentes e repetíveis para carregar dados operacionais (ETL) que suportam as operações do dia a dia. Além disso, o armazenamento de dados traduz as informações heterogêneas em modelos comuns, nomes e definições e fornece um meio de disponibilizar a informação consistente à tomada de decisão. Uma base de dados bem estruturada apresenta informações de acordo com temas específicos, integra dados de múltiplas fontes, armazena dados históricos que permanecem consistentes. Independentemente do momento em que é consultada, os dados guardados no repositório não irão mudar (Nussbaumer, Revels, 2013).

Em conclusão, o conceito de DW (Kimball, Ross, 2013) é um dos tópicos, nos dias de hoje, mais interessantes da indústria da computação. Para os executivos de empresas, este conceito alberga vantagem competitiva e significativa para as suas empresas, enquanto os gestores de sistemas de informação encaram este conceito como uma maneira eficaz de superar os obstáculos tradicionais para fornecer informações comerciais para outros gestores e utilizadores finais.

A Figura 3 reflete a relação entre os componentes de um sistema de DW, uma vez que esta faz uso constante de ferramentas que realizam tarefas definidas. Desta forma, através da imagem, é possível verificar a existência de: camada de *staging* que corresponde às camadas de bases de dados operacionais e fontes de dados externas e internas do DW; um conjunto de estruturas de dados analíticos, ou seja, o próprio DW; um sistema de gestão de base de dados (SGBD) configurado especialmente para responder aos requisitos analíticos dos sistemas de DW; um componente que disponibiliza aos utilizadores finais o acesso aos dados do DW e, por último, um repositório para armazenar e gerir os metadados que ajuda os utilizadores na recolha e interpretação de dados, para a criação de relatórios de apoio à decisão, por exemplo.



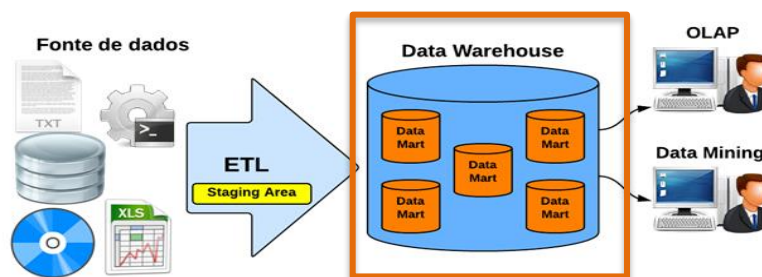


Figura 3: Relação entre os diferentes componentes de um sistema de DW

Fonte: Disponível em: <<https://canaltech.com.br/business-intelligence/conhecendo-a-arquitetura-de-data-warehouse-19266/>>  
Acesso em ago. 2018.

### 2.3.2 Modelação Dimensional

O modelo dimensional é uma técnica utilizada atualmente no desenho de DW. Segundo Ralph Kimball (2013), o modelo dimensional é a técnica mais viável para entregar dados aos utilizadores do DW porque apresenta os dados de uma forma simples e intuitiva, além de permitir um elevado desempenho nas consultas, uma vez que fornece dados compreensíveis aos utilizadores e permite a entrega rápida de consultas SQL (*Queries*).

A modelação dimensional (Kimball, Ross, 2013) constitui, desta forma, a primeira parte de organização de um DW e envolve quatro importantes etapas em que os dados são modelados em tabelas dimensionais ligadas a uma tabela de Factos e estas relações são representadas num esquema em estrela, que serve como base para a criação dos relatórios de apoio à decisão:

- 1 – Identificar o processo de negócio a modelar;
- 2 – Declarar o grão da tabela de Factos, onde se determina o nível máximo de detalhe (fino ou grosso), se identificam as dimensões e se determina o detalhe a guardar nas dimensões, definindo o significado associado a cada uma linha da tabela de Factos;
- 3 – Modelar as dimensões do negócio em tabelas, incluindo hierarquias de dados, tendo em conta o processo de negócio escolhido e com detalhe adequado ao grão;
- 4 – Identificar medidas numéricas na tabela de Factos.

O DW (Kimball, Ross, 2013) é um repositório de dados multidimensionais composto por:

- Medida (substantivo): valor numérico usado para avaliar um processo de negócio, isto é, são atributos que têm representatividade de desempenho como indicador de negócio.

- Dimensão (substantivo): entidade independente que agrupa dados de negócio e que participa na análise de medidas que pode ser uma data, por exemplo. Para identificar cada linha das tabelas de dimensões, temos as chaves substitutas (*surrogate keys* que constituem a *primary key* da dimensão) de pesquisa eficiente e cujo objetivo é permitir a junção da tabela de dimensão com a tabela de Factos. As tabelas de dimensão integram um conjunto diversificado de atributos, ou colunas, pelos quais os processos de negócio considerados na tabela de Factos podem ser analisados. Estes atributos integram, habitualmente, descrições que permitem contextualizar as métricas em análise.

- Facto (verbo): tabela primária num modelo dimensional que corresponde a um evento expresso através de dimensões e medidas dos processos de negócio. As tabelas de Factos possuem chaves estrangeiras (*Foreign Keys*) que referenciam as tabelas das dimensões e medidas numéricas ou métricas sobre o negócio que são dados numéricos. São tipicamente grandes, ocupam 90% da dimensão total do DW, não apresentam dados redundantes, são estreitas, com poucas colunas e precisam de atualização com alguma frequência, sobretudo na parte de inserção dos dados.

Existem três tipos de tabelas de Factos: tabela de Factos transacional, tabela de Factos instantâneos periódicos (*periodic snapshots*) e tabela de Factos instantâneos cumulativos (*accumulation snapshots*). Por ser abordada neste projeto, será referenciada a tabela de Factos transacional. Esta tabela é a mais comum dos três tipos e é uma tabela que regista uma linha por cada transação, ou seja, uma linha é criada sempre que ocorre um evento. Normalmente, são as tabelas que contém dados no nível mais detalhado e, por isso, ocupam mais espaço em disco.

Os modelos de dados que recorrem a técnicas de relacionamento entre entidades são inadequados para modelos de DW. Partindo deste princípio, é necessário recorrer a uma modelação multidimensional para conceber um modelo de dados para este tipo de sistemas. Segundo Ramos e Santos (2009), este tipo de modelação traduz-se num modelo de dados fácil de compreender e utilizar e, ao mesmo tempo, proporciona uma melhor otimização do sistema. A modelação multidimensional apresenta como componentes básicos os Factos, as dimensões e as hierarquias (Golfarelli, Maio, Rizzi, 1998), e pode ser implementada sobre um dos seguinte três esquemas (Chaudhuri, Dayal, Narasayya, 2011): Esquema em Estrela (*Star Schema*); Esquema em Constelação (*Constellation Schema*); Esquema em Floco de Neve (*Snowflake Schema*). Por ter relevância neste projeto, só será explicado o *Star Schema*.

O *Star schema* (exemplo na Figura 4) foi criado por Ralph Kimball (Kimball, Ross, 2013), para sistemas de apoio à decisão, configurando uma técnica de modelação dimensional. Cada processo de negócio é representado por um modelo que consiste numa tabela de Factos que contém métricas de um processo de negócio, rodeado por tabelas de dimensões que englobam o contexto. Tem características estruturais de uma estrela, adquirindo frequentemente a denominação de *Star join* (Kimball & Ross, 2013) em que é constituído por uma única tabela de Factos e por várias tabelas de dimensões.

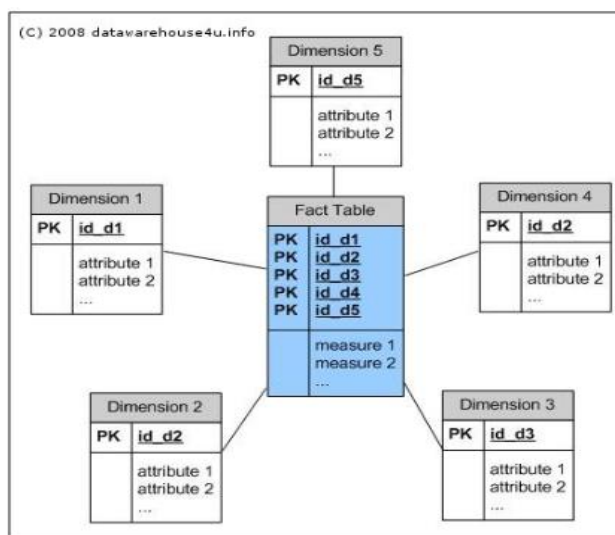


Figura 4: *Star Schema*

Fonte: Disponível em: < <http://Datawarehouse4u.info/Data-warehouse-schema-architecture-star-schema.html> > Acesso em ago. 2018.

Na modelação multidimensional deve, também, ter-se em consideração os requisitos do negócio para os dados que se encontram nas tabelas de facto e dimensões, especialmente quando estas são atualizadas. Ao contrário da Tabela de Factos que, por definição, representa uma série temporal de medições com uma dimensão temporal explícita (Nemec, Nguyen, Tjoa, Windisch, 2006), as tabelas de dimensão requerem um tratamento especial. As tabelas de dimensão não sofrem alterações de uma forma previsível pois evoluem lentamente à medida que o tempo passa (Rainardi, 2008), e, normalmente quando sofrem alterações são mudanças físicas (por exemplo, nova morada de cliente) ou correções de erros nos dados. Este aspeto refere-se à necessidade de efetuar alterações nos atributos de uma dimensão, pois estes podem sofrer alterações ao longo do tempo e os respetivos registos necessitarem de ser alterados. Existem três técnicas que permitem registar mudanças lentas nas dimensões:

- A técnica do tipo 1, que consiste na sobreposição do valor. Nesta técnica, o valor antigo do atributo de cada linha da dimensão é sobreposto pelo valor corrente. Neste caso, o atributo reflete sempre a atribuição mais recente. A vantagem do tipo 1 consiste na rapidez, ou seja, nas tabelas de dimensão, o valor preexistente é sobreposto pelo valor corrente e as tabelas de Factos mantêm-se sem qualquer tipo de alteração. O único problema consiste na manutenção do histórico de alterações dos atributos, ou seja, ao existir sobreposição de valores apenas são armazenados no DW os valores recentes que sobrepõem os anteriores;
- A técnica do tipo 2, que consiste em adicionar uma linha na dimensão, permitindo armazenar o histórico. Esta técnica segmenta perfeitamente o histórico da tabela de Factos;
- Por fim, a técnica do tipo 3, em que se adiciona uma coluna à dimensão, a qual não é abordada neste projeto.

A modelação num sistema de bases de dados multidimensionais permite análises profundas e o estabelecimento de relações de dados que não são atingidos no modo convencional.

### 2.3.3 Processos de ETL

Analisados os conceitos relacionados com as arquiteturas dos DW e das estruturas dos modelos de dados, descrevem-se de seguida os mecanismos que permitem o povoamento dos repositórios. Como referido anteriormente, um DW é um repositório de dados, preparado especialmente para ser utilizado para apoiar a tomada de decisão nas organizações. Para que seja efetivamente possível, os dados devem ser extraídos dos sistemas OLTP, devem ser transformados e só depois carregados para o DW (Ariyachandrea, Watson, 2010). A este processo dá-se o nome de ETL (MicroStrategy University, 2008) e inclui um conjunto de ferramentas especializadas de extração de dados das fontes de dados (armazenar informações sobre a estrutura do sistema de origem e conteúdo), transformação dos dados na DSA e carregamento para a DPA, que permitem tratar a complexidade encontrada nos dados, tratando da sua homogeneidade, a sua posterior limpeza e o respetivo carregamento para o DW (Simitsis, Skiadopoulos, Vassiliadis, 2002).

Os processos de ETL ocupam até 80% do esforço em projetos de BI, de modo que o seu elevado desempenho é vital para que seja capaz de processar elevadas quantidades de dados e para manter uma base de dados atualizada.

A etapa da extração corresponde ao processo de obtenção de uma cópia dos dados que estão nas fontes de dados operacionais em diversos formatos e estruturas, sendo importante conhecer a validade dos dados. Nesta etapa de extração, poder-se-ão copiar todos os dados, ou só aqueles que foram modificados desde a última extração. A responsabilidade envolvida nesta etapa é a receção dos dados das várias fontes. As extrações dos dados podem ser distinguidas, numa primeira extração, que corresponde ao primeiro carregamento dos dados para o DW vazio, e as posteriores extrações incrementais dos dados que vão atualizar DW com os novos dados que vão surgindo gradualmente, da atividade diária da organização ao longo do tempo (Ramos, Santos, 2009).

A etapa da transformação, facilmente dispensável nalguns casos e muito importante noutros, consiste em converter os dados em informação. A manipulação dos dados (duplicação de dados, diferentes representações de valor, dados com erros e dados em falta) nesta etapa exige, por vezes tarefas mais complexas como: limpeza dos dados (transformação de “masculino” para “M” no DW); tradução livre de valores codificados (transformar “masculino” e “Sr.” em “M”); aplicação da transformação apenas a determinadas categorias de linhas e/ou colunas (transposição ou rotação); fusão (*merge*) ou agregação dos dados de fontes distintas; resumo de várias linhas de dados; tratamento de exceções; eliminação de duplicados; junção de dados cuja origem é de várias fontes diferentes; separação de uma coluna em várias. É, também, nesta fase onde são geradas as tabelas de dimensões e as tabelas de Factos, incluindo as respetivas chaves substitutas (para as tabelas de dimensões) e estrangeiras (para as tabelas de Factos).

Por último, a etapa do carregamento dos dados para uma DW ou *Data Marts* corresponde à última etapa do processo ETL e é realizada aquando da finalização do tratamento dos dados das tabelas de dimensões e após a criação das tabelas relacionais para serem inseridas na base de dados relacional. De uma forma geral, carregamento consiste em povoar, habitualmente através da inserção de registos, as tabelas de destino no DW, que estão organizadas de acordo com o modelo de DW escolhido. O carregamento pode ser simples (reescrever dados novos por cima de antigos) ou mais completo em termos de dados históricos (mantendo um registo de todas as alterações efetuadas).

Resumindo, nos atuais sistemas de ETL, o fluxo de dados é praticamente unidirecional, a partir das fontes de dados para o DW (Castellanos, Dayal, Simitsis, Wilkinson, 2009).

A *framework* dos processos de ETL, mostrado na Figura 5, consiste no caminho que os dados seguem até estarem de acordo com a estrutura do DW. Os dados são extraídos de diferentes fontes de dados (internas ou externas), posteriormente serão carregados para a DSA onde serão transformados e limpos, antes de serem carregados para o DW. De realçar que os processos ETL são fundamentais na construção e manutenção dos sistemas de DW (Shaker et. al., 2011), pois é através destes processos que as empresas estruturam os dados de maneira a que futuramente possam ser analisados e entregues ao utilizador final.

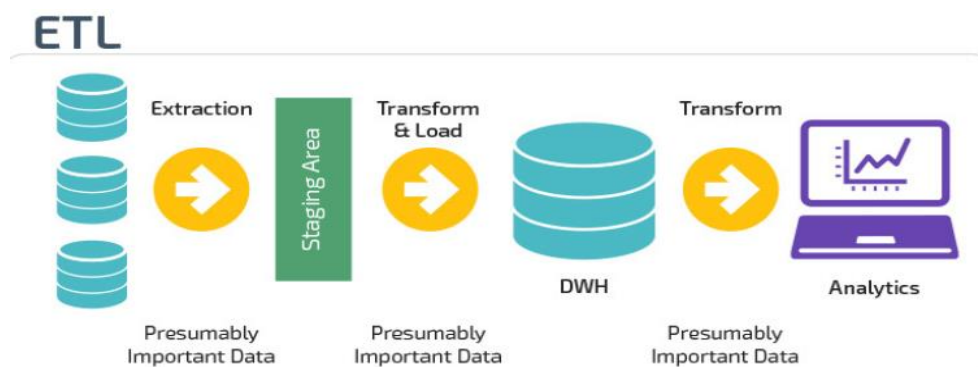


Figura 5: Coordenação entre as diferentes fases do Processo ETL

Fonte: Disponível em: <<https://panoply.io/data-warehouse-guide/3-ways-to-build-an-etl-process/>> Acesso em ago. 2018.

## 2.3.4 Microsoft Power BI

Um dos principais aspetos do BI é o de proporcionar ferramentas de fácil utilização, que faz com que esteja verdadeiramente disponível para o utilizador final. (Azevedo, Santos, 2009). O uso da gestão do conhecimento, da gestão de conteúdo e de plataformas organizacionais é a solução para partilhar informações num ambiente empresarial, admitindo que se pretende trabalhar em conjunto com a finalidade de atingir os objetivos comuns da empresa (Pant, 2009). Atualmente, a entrega e a partilha de informação já está disponível em versões *mobile*, para telemóveis e *tablets*. É o caso do Microsoft Power BI, uma aplicação da Microsoft que comporta inúmeras ferramentas de análise de dados referentes a um determinado negócio. Fazem parte do mundo *Power BI*, inúmeras aplicações com métricas atualizadas em tempo real, e disponíveis em todos os dispositivos, que podem ser acedidas pelos vários utilizadores de uma organização. O *Power BI* consegue albergar múltiplos dados de uma organização, guardando-os na *Cloud* ou localmente. Recorrendo a *gateways* do *Power BI* é ainda possível a ligação às bases de dados que estejam no *SQL Server*, modelos de *Analysis Server* ou outras fontes de dados. É através da partilha e uso da informação adquirida com utilização de ferramentas de BI que a empresa irá tirar vantagens da sua estrutura, com decisões apoiadas em factos e numa análise mais rigorosa da empresa e da sua *performance*.

### 2.3.4.1 Dashboards em Business Intelligence

Os *dashboards* são uma funcionalidade do serviço *Power BI* e não estão disponíveis na ferramenta *Power BI Desktop*. *Dashboard* (Microsoft Docs, Dashboards no serviço Power BI, 2017) é entendido como um painel interativo, numa única página, muitas vezes designada de tela, que usa visualizações para apresentar uma única situação de análise sobre o desempenho dos processos de negócio de uma organização. Como está limitado a uma página, contém apenas os elementos mais importantes dessa análise. As visualizações de um *dashboard* são chamadas de mosaicos, atualizáveis à medida que os dados vão sendo alterados e fixos ao *dashboard* a partir do painel de relatórios, que se baseiam num conjunto de dados. Estas visualizações podem ser provenientes de um ou vários conjuntos de dados ou relatórios. Um *dashboard* combina dados no local e na *Cloud* ao fornecer uma vista consolidada, independentemente de onde os dados residem.

Os *dashboards* podem ser representados num painel virtual composto por uma ou mais camadas de instrumentos virtuais como gráficos de barras, colunas, bolhas, entre outros. Podem ainda ser

identificados três tipos de *dashboards*: operacionais (acompanham e monitorizam os processos de negócio e são da responsabilidade de supervisores e trabalhadores), táticos (acompanham os processos departamentais e são da responsabilidade de gestores e analistas) e estratégicos (acompanham a execução dos objetivos pré-definidos e são da responsabilidade dos gestores, executivos e restantes elementos).

Um *dashboard* permite juntar todo o tipo de relatórios num único documento e, normalmente, exibe apenas a informação mais importante recolhida desse conjunto de dados, existindo a possibilidade de destacar um determinado *dashboard*. Não são tão detalhados como os relatórios, mas podem dar respostas rápidas às perguntas mais importantes. É um painel de informação que oferece aos utilizadores um alto nível de interatividade na maneira como os dados são apresentados, permitindo juntar e combinar vários componentes (tabelas, gráficos, entre outros) e explorar/manipular os dados utilizando as várias capacidades OLAP como o *drill* e o *slice and dice*. Dependendo do papel no projeto, distinguem-se dois tipos de utilizadores: os criadores de *dashboards* e os consumidores de *dashboards*. Os primeiros criam *dashboards* para o seu próprio uso ou para outros enquanto que os consumidores estão aptos a consultar e a receber *dashboards* de terceiros.

De facto, um *dashboard* é altamente interativo e personalizável e funciona como uma entrada num relatório de um conjunto de dados e uma excelente maneira de monitorizar o negócio, procurar respostas e ver todas as métricas importantes para o negócio.

#### 2.3.4.2 Relatórios em *Business Intelligence*

Os relatórios de BI (Microsoft Docs, *Relatórios no Power BI*, 2017) são uma funcionalidade do serviço Power BI e da ferramenta *Power BI Desktop* e constituem vistas particulares de várias perspetivas sobre um único conjunto de dados, com diferentes visualizações, construídos para satisfazer uma determinada necessidade de acordo com períodos de tempo específicos ou elementos exclusivos de um projeto.

Um relatório pode ter uma única visualização ou páginas cheias de visualizações, cada uma, representando um grupo de informações e, é constituído por elementos como tabelas de dados, *dashboards*, cubos de dados, análises de *Data Mining*, entre outros. Dependendo do papel desempenhado no projeto, uma pessoa pode criar relatórios para o seu próprio uso ou para partilha (criador do relatório) ou ser alguém que possa visualizar, usar e receber relatórios de outros (consumidor do relatório).

Nos relatórios, as visualizações não são estáticas, sendo possível adicionar e remover dados, alterar os tipos de visualização, aplicar filtros e segmentações para descobrir informações e procurar respostas. Um relatório é, no fundo semelhante a um painel, mas mais, um relatório é altamente interativo e altamente personalizável e a atualização de visualizações é possível à medida que os dados subjacentes mudam. Os principais objetivos da construção de relatórios consistem na disponibilização de informação atual e detalhada ao utilizador, uma vez que os dados estão no menor nível de granularidade e flexibilidade para permitir que os utilizadores finais possam criar as suas próprias visualizações.

Concluindo, os relatórios de apoio à decisão ajudam as empresas na tomada de decisão, permitindo a sua autoavaliação e avaliação por entidades externas. São também eles criados para que os utilizadores finais consigam aceder aos dados, uma vez que todas as camadas anteriormente descritas são invisíveis ao cliente final.

### 2.3.4.3 Apps Workspaces e Apps do Microsoft Power BI

O serviço de Power BI permite duas funcionalidades através da utilização dos dois componentes das Apps Power BI: o *App Workspace* e a *App* propriamente dita.

As *App Workspaces* (Microsoft Docs, Collaborate in your Power BI app workspace, 2017) são áreas de trabalho onde podem ser criados e/ou editados relatórios e *dashboards* para um número variado de colaboradores. Tem inúmeras funções, das quais as mais importantes são as seguintes: possui um vínculo ao espaço de trabalho da aplicação, por ter a opção de publicar relatórios ou *dashboards*; sugere um ambiente de preparação para relatórios e *dashboards* e gere os utilizadores da aplicação (*App*). Sobre este *Workspace*, é construída a *App* propriamente dita, a qual é responsável por disponibilizar o conteúdo pretendido para um determinado conjunto de colaboradores.

### 2.3.4.4 Self-Service BI

O conceito inerente de *self-service* BI do *Power BI* (Gallo, 2018) traduz-se na criação, pela comunidade de utilizadores de negócios, das suas próprias análises, sem necessitarem de recorrer a equipas especializadas de TIs para a criação de relatórios. Assim, torna-se mais rentável apostar na formação de colaboradores da empresa, dando assistência e criando documentação, por forma a que os utilizadores consigam utilizar eficazmente as ferramentas e os dados disponibilizados. Para ser possível criar um sistema de *self-service* BI, a aplicação disponibilizada aos utilizadores deve permitir que os utilizadores criem os seus próprios relatórios e explorem a informação no DW sem precisarem de conhecer a sua estrutura. A interface gráfica desta aplicação deve ser fácil de utilizar e perceber.

## 2.4 Tipos de Arquitetura: *On-prem*, *Cloud* e Híbrido

Com o avanço tecnológico dos últimos anos, cada vez mais as empresas precisam de se adaptar e atualizar face à concorrência. Diante deste cenário e devido aos altos custos dos equipamentos de tecnologia necessários para montar um *data center* próprio, manutenção de licenças, *hardware*, e contratação de profissionais especializados, é cada vez mais importante integrar e gerir sistemas.

Existem três modelos para definir a arquitetura da solução (Figura 6) (Brunton, 2015): solução *on-premises*, solução na *Cloud* (Microsoft Azure) e solução híbrida (*on-premises* e *Cloud*). Cada uma foi criada para atender às respetivas necessidades de cada empresa. A solução *on-premises*, utilizada neste projeto, é uma solução que é implementada na organização, utilizada por empresas com grande espaço físico para manter uma infraestrutura de *data center* (Cliente X.) e uma grande disponibilidade económica para investimento em *hardware* e *software*, principalmente para o seu licenciamento. Para implementar uma solução deste nível é necessário ter em conta diversos aspetos: existência de servidores e sistemas de virtualização; suporte económico e manutenção dos vários produtos envolvidos; infraestruturas de comunicação locais e ocupação de *data center*; pessoal técnico com conhecimento para criar a configuração da solução de raiz, especialista em administração de sistemas para a gestão, monitorização e resolução de problemas; e, outras encargos extra como energia elétrica, por exemplo; sistemas complementares para *backup* da informação. É necessário ter cuidado com a implementação desta solução uma vez que a escalabilidade da solução é dispendiosa e difícil, devido à necessidade de equipas profissionais especializadas e de equipamentos que necessitam de ser configurados; necessidade

de fazer *backup* que exige um serviço na *Cloud* ou que os dados sejam replicados noutro ponto geográfico; e, morosidade e custos das atividades de configuração, reestruturação e redimensionamento da solução.

A solução na *Cloud* é uma opção mais personalizada e construída sob medida para atender às características de determinada empresa que não pode suportar essa infraestrutura dentro das suas dependências. Ou seja, os recursos utilizados, processamento e armazenamento ocorrem remotamente em equipamentos da entidade que presta um determinado serviço a uma empresa. Neste cenário, a empresa contratada é responsável por proteger os dados hospedados, pela manutenção de *hardware* e *software*, e o cliente apenas paga pelo serviço *hardware* e *software* que utilizar. Na implementação de uma solução na *Cloud*, é necessário criar uma conta no Microsoft Azure, criar máquinas virtuais quando necessário, sendo que estas já se encontram pré-configuradas, acrescentando apenas algumas configurações extra que sejam necessárias e utilizar as ferramentas disponibilizadas pelo serviço.

Por último, a solução híbrida é uma opção que tenta conciliar ambos os modelos descritos anteriormente (parte da solução na organização e outra parte na *Cloud*, no Microsoft Azure), equilibrando os gastos operacionais e determinando, essencialmente, o que pode estar ou não em ambiente público ou privado.

Independente da escolha, todos os modelos citados atendem aos níveis aceitáveis de segurança e a importância de integrar e gerir modelos é poder ter maior escalabilidade operacional acompanhando o crescimento ou a redução da prestação das empresas, por forma a adequar-se ao cenário atual, com rapidez no tempo de implementação, sem perda de informação, e proporcionando um baixo custo operacional.

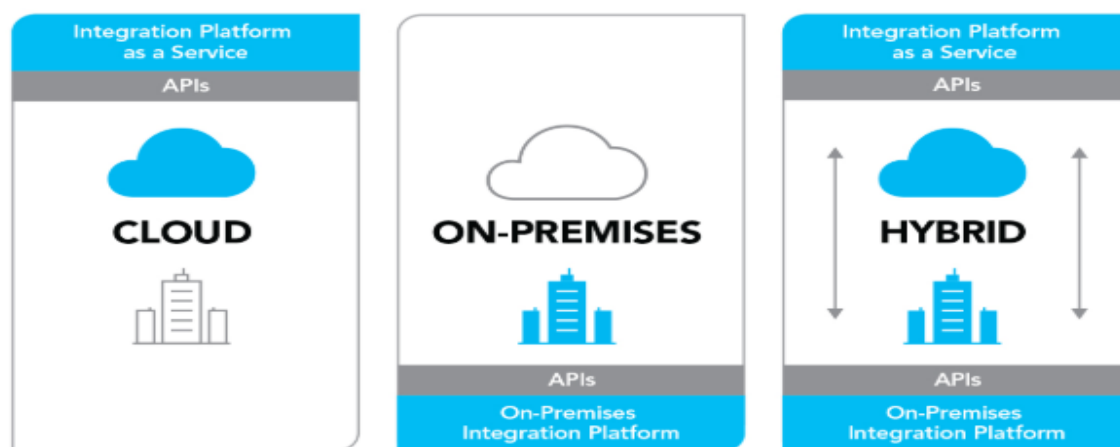


Figura 6: Tipos de Arquitetura

Fonte: Disponível em: < [https://www.softwareag.com/fr/products/webmethods\\_integration/Cloud\\_integration/default](https://www.softwareag.com/fr/products/webmethods_integration/Cloud_integration/default)>  
Acesso em ago. 2018.

## 2.5 Sumário

Neste capítulo, foram descritos os conhecimentos das regras de negócio e melhores práticas na implementação de soluções de BI, noções de OLAP e OLTP e, também demonstrado o conhecimento necessário e especializado de arquiteturas de sistemas de informação para sustentar a extração e normalização da informação a partir dos sistemas operacionais, sua vertente histórica de evolução e sua



vertente tecnológica. Seguidamente, detalhou-se o conceito de DW, Modelação Dimensional, ETL, onde se especificaram as suas fases e principais funcionalidades. Por último, foram também apresentados os conceitos do universo Microsoft *Power BI*, nomeadamente, diferença entre *dashboards* e relatórios em Microsoft *Power BI*, noções de App, App *Workspace* e *Self-Service BI* do *Power BI*. Terminou-se o capítulo com a distinção entre os três tipos de arquitetura possíveis de implementar num projeto de BI.



## Capítulo 3

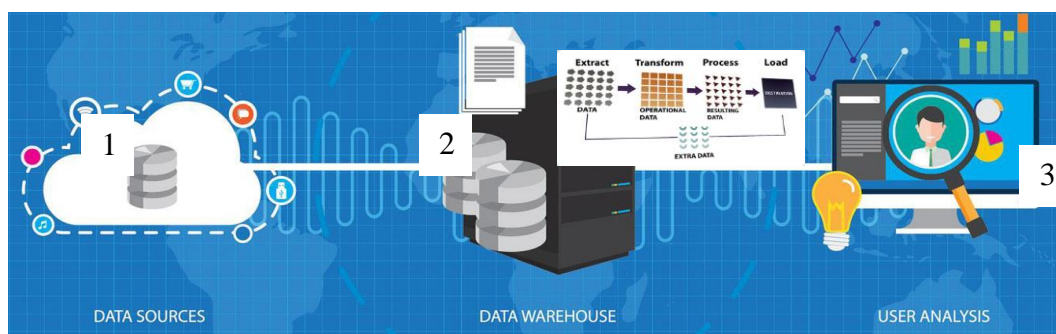
### Trabalho Realizado

Neste capítulo, são apresentados os métodos de trabalho seguidos, as ferramentas utilizadas, as tecnologias, os conhecimentos que contribuíram para a construção do produto final deste projeto e os relatórios em *Power BI*, que respondem às exigências do relatório mensal. É ainda realizada uma descrição dos detalhes teóricos, práticos e técnicos que permitem a compreensão da abordagem adotada na execução deste projeto bem como a resolução de problemas que foram surgindo. Ainda neste capítulo, se apresenta o planeamento do projeto e uma análise dos desvios ao plano de trabalho.

#### 3.1 Ambiente de Trabalho

Neste tópico, são abordadas as tecnologias, as ferramentas e as linguagens utilizadas, descrevendo o que são, para que servem e como se enquadram em cada fase do projeto desenvolvido. É ainda mencionado neste capítulo: o processo de desenvolvimento de *software*, exposição das diversas fases do ciclo de vida do projeto, o trabalho em si realizado, organização do trabalho de equipa e respetivo fluxo de trabalho. Do mesmo modo, é dada, ainda, uma visão geral da arquitetura e suas funcionalidades bem como algumas considerações quando necessárias.

Durante o desenvolvimento do projeto, a autora deu apoio noutros projetos que iam de encontro ao tema do trabalho de projeto, sendo que a implementação de um *reporting* corporativo assume maior destaque (Figura 7), por ser mais complexo e por acompanhar a implementação de um sistema de BI desde a raiz até à sua disponibilização para o utilizador final, o Cliente X.



1. Análise dos dados e requisitos por meio de uma prova de conceito com os relatórios mensais do Cliente X.
2. Modelação do *Datawarehouse* e Desenho e Implementação dos processos de ETL (dimensões e Factos).
3. Camada de Visualização: Construção dos Relatórios em Microsoft *Power BI* para cada fonte de dados.

Figura 7: *Workflow* de Implementação do DW Corporativo

Fonte: Adaptada de Figura disponível em: < <http://computecheexplorer.com/data-warehouse-training-courses-pune.php> >  
Acesso em ago. 2018.

### 3.1.1 Ferramentas, Linguagens e Tecnologias utilizadas

Depois de discutir os conceitos associados aos sistemas de BI (Capítulo 2), importa agora referir as ferramentas e tecnologias que suportam a sua criação e gestão. Estas aplicações, infraestruturas, ferramentas e tecnologias são, sumariamente, descritas em seguida e, trazem benefícios estratégicos para as organizações, ajudando-as no suporte de decisões e previsões de negócio.

Durante a realização do estágio que conduziu à criação e produção deste trabalho de projeto, as tecnologias utilizadas foram: SSIS do Microsoft SQL Server e Microsoft Power BI. Como ferramentas utilizou-se o Power BI Desktop, o Microsoft Excel e o MDS (para a construção de tabelas de mapeamento de dados). As ferramentas identificadas foram instaladas e configuradas. Como linguagens, foram utilizadas as seguintes: linguagem SQL para a construção das tabelas, *views*, *stored procedures* e *jobs* e a, linguagem DAX na construção de métricas, no Power BI. A seleção das tecnologias para o desenvolvimento foi condicionada pelo licenciamento existente no Cliente X.

#### 3.1.1.1 Microsoft Excel

O Microsoft Office Excel é um editor de folhas de cálculo produzido pela Microsoft que permite a pesquisa de informação que se encontra armazenada. Os seus recursos incluem uma interface intuitiva e capacidade de ferramentas de cálculo e construção de gráficos, que tornam o Microsoft Excel, uma das maiores e mais usadas aplicações para computador.

Esta ferramenta foi importante para a análise, levantamento de requisitos e implementação dos modelos de dados que serviu de suporte no estudo dos principais indicadores de negócio (objetivo 1 do projeto – ver ponto 1.4.1 , página 8).

#### 3.1.1.2 SQL Server Integration Services

O Microsoft SQL Server Management Studio (SSMS) [Microsoft Docs, *SQL Server Management Studio (SSMS)*, 2017] constitui um ambiente integrado (*client tool*) que permite o acesso, a configuração, a gestão e a administração dos vários componentes do Microsoft SQL Server, possibilitando, este último, o armazenamento de grandes quantidades de dados de diversas fontes.

Por sua vez, o SSMS fornece ferramentas para configurar, monitorizar e administrar instâncias de SQL. Os usos de SSMS englobam: implementar, monitorizar e atualizar os componentes de camada de dados usados pelas aplicações, bem como criar consultas e *scripts*. Esta ferramenta está diretamente ligada aos objetivos 2, 3, 4 e 5 (ver ponto 1.4.1 , página 8), sendo responsável por guardar os dados e procedimentos necessários para alimentar os relatórios de apoio de decisão. No SQL Server, foram criadas as bases de dados, tabelas, *views*, *stored procedures* (SP) e *jobs* das três fontes de dados operacionais, canalizando-os para os respetivos modelos multidimensionais de dados, posteriormente carregados para um DW.

O SQL Server possui, ainda, um leque de variadíssimas ferramentas, tais como, extração de dados (ETL), ferramentas de relatórios (SQL Server Reporting Services – SSRS), ferramentas de análise (SQL Server Analysis Services – SSAS) e funcionalidades que permitem criar aplicações analíticas tornando-se numa plataforma de BI completa. O SQL Server Integration Services (Microsoft Docs, *SQL Server Integration Services*, 2017) é um componente do *software* de bases de dados do Microsoft SQL

*Server* que pode ser usado para executar uma ampla variedade de tarefas de migração de dados. O *Microsoft Integration Services* é uma plataforma para a construção de soluções de integração de dados a nível empresarial. Este é usado para resolver problemas comerciais complexos, copiando ou descarregando arquivos, enviando mensagens de *e-mail* em resposta a eventos, atualizando repositórios de dados, limpando dados e gerindo objetos e dados do *SQL Server*. Os pacotes podem ser usados sozinhos ou em conjunto com outros pacotes para atender às necessidades complexas do negócio. O *Integration Services* pode extrair e transformar dados de uma grande variedade de fontes, como arquivos de dados *eXtensible Markup Language* (XML) e fontes de dados relacionais e, depois carregar os dados num ou mais destinos. Inclui, ainda, um conjunto de tarefas, transformações integradas e ferramentas para a construção, execução e gestão de pacotes. As ferramentas gráficas do *Integration Services* podem ser utilizadas para criar soluções sem utilizar uma única linha de código ou até mesmo programar o extenso modelo de objeto do *Integration Services* para criar pacotes e codificar tarefas personalizadas. Esta ferramenta foi responsável por implementar os processos de ETL que constituiu o objetivo 2 do projeto (ver ponto 1.4.1 , página 8).

### 3.1.1.3 Microsoft Power BI: *Power BI Desktop*

Microsoft *Power BI* (Microsoft Docs, O que é o Power BI?, 2017) constitui um conjunto de ferramentas de análise de negócios que fornece informações para toda a organização e serviços para criar, implementar e gerir relatórios, para além dos recursos de programação que permitem estender e personalizar a funcionalidade dos relatórios gerados com esta ferramenta. O *Power BI* fornece serviços de BI baseados na *Cloud*, juntamente com uma interface baseada em *desktop*, chamada "*Power BI Desktop*". Em março de 2016, a Microsoft lançou um serviço adicional chamado *Power BI.com* na sua plataforma de nuvem, *Azure*. A diferença principal do produto é a capacidade de carregar visualizações personalizadas.

O *Power BI Desktop* (Microsoft Docs, O que é o Power BI Desktop?, 2017) foi a ferramenta escolhida para a visualização dos dados, pois permite a construção de tabelas e gráficos dinâmicos e filtragens com grandes volumes de dados, que rapidamente ajudam na tomada de decisão. Esta plataforma tem uma interface gráfica e interativa, sendo mais fácil os utilizadores explorarem os dados através de operações como o *drill* e experimentar diferentes tipos de apresentações como mapas, gráficos ou animações. Esta aplicação da Microsoft *Power BI*, pode ser instalada localmente, permitindo a transformação e a visualização de dados e, constitui uma ótima ferramenta de criação de relatórios. O uso desta ferramenta no projeto consistiu na criação de relatórios automáticos e detalhados com base nos dados preparados que provinham das três fontes (objetivo 6 – ver ponto 1.4.1 , página 8). Ainda no contexto da tecnologia Microsoft *Power BI*, foi usada a linguagem DAX para facilitar o cálculo de algumas métricas associadas a campos dos relatórios operacionais desenvolvidos. Através do uso desta tecnologia, foi possível incutir o conceito de *self-service* BI (objetivo 7 – ver ponto 1.4.1 , página 8) que permite aos utilizadores finais uma maior flexibilidade nas análises, maior autonomia no acesso à informação proveniente das várias fontes de dados, maior facilidade em explorar grandes quantidades de dados, observá-los de diferentes formas e encontrar respostas. Na construção dos relatórios, foi usado o *Power BI Pro*, daí ter sido necessário pedir licenças de *Power BI* para os utilizadores.

### 3.1.1.4 Linguagem SQL

A linguagem SQL (Juliano, 2008) é uma linguagem declarativa padrão que é usada para definir, questionar, armazenar, manipular e recuperar dados em bases de dados. É uma linguagem específica de domínio usada na programação e projetada para gerir dados mantidos num *Relational Database Management System* (RDBMS) ou para o processamento num fluxo de dados relacionais (*Relational Data Stream Management System* – RDSMS). Em comparação com *Applications Program Interface* (APIs) antigas de leitura, como *Indexed Sequential Access Method* (ISAM) ou *Virtual Storage Access Method* (VSAM), o SQL oferece duas vantagens principais: primeiro, introduz o conceito de aceder a muitos registos com um único comando e, em segundo lugar, elimina a necessidade de especificar como chegar a um registo, com ou sem índice. Originalmente baseada em álgebra relacional, a linguagem SQL consiste em vários tipos de afirmações, que pode ser classificada de forma informal como sublinguagens, comumente: uma linguagem de consulta de dados (*Data Query Language* – DQL), uma linguagem de definição de dados (*Data Definition Language* – DDL), uma linguagem de controlo de dados (*Data Control Language* – DCL) e uma linguagem de manipulação de dados (*Data Manipulation Language* – DML).

Esta linguagem foi a linguagem usada no SQL Server para a criação de tabelas, *views*, *stored procedures* e implementação de *jobs* que facilitaram a execução dos processos (objetivos 2, 3, 4 e 5 – ver ponto 1.4.1 , página 8).

### 3.1.1.5 Linguagem DAX

A linguagem DAX é uma nova linguagem de fórmulas (Microsoft Docs, Descrição Geral da Linguagem DAX (*Data Analysis Expressions*), 2018) que permite aos utilizadores definir cálculos personalizados em tabelas de *PowerPivot* (colunas calculadas) e em tabelas dinâmicas do Microsoft Excel (medidas). Através desta linguagem, é possível criar fórmulas nas camadas de dados, como colunas calculadas e medidas bem como agrupar os dados.

O DAX inclui algumas das funções utilizadas em fórmulas do Microsoft Excel e funções adicionais concebidas para trabalhar com dados relacionais e efetuar a agregação dinâmica. As fórmulas do DAX são muito semelhantes às fórmulas do Excel. De uma forma mais simples, o DAX (objetivo 6 – ver ponto 1.4.1 , página 8) ajuda o utilizador a criar nova informação a partir de dados já existentes no seu modelo, em Microsoft *Power BI*.

### 3.1.1.6 Master Data Services

O *Master Data Services* (MDS) é uma solução existente em várias versões do SQL Server (Microsoft Docs, Master Data Services, 2014) que permite a gestão de dados mestre (*Master Data Management* – MDM). Esta gestão acontece quando uma determinada organização possui uma lista de dados não transacionais (dados que possuem, normalmente, uma data inicial e uma data final e que, por isso, definem quando a condição ou o valor dos dados permanece inalterado, comprometendo o desempenho da organização, como por exemplo, estados de um determinado pedido da fonte de dados A), os quais necessitam de fazer parte de listas atualizáveis frequentemente. Um projeto MDM geralmente inclui uma avaliação e reestruturação dos processos internos de negócios, juntamente com a implementação da tecnologia MDM. O resultado de uma solução de MDM bem-sucedida é a confiabilidade dos dados centralizados que podem ser analisados, resultando em melhores decisões de

negócios. No âmbito do projeto, a solução de MDS foi utilizada para a construção de tabelas de mapeamento, para que alguns campos das fontes de dados fossem descodificados. Esta solução permitiu a centralização dos dados e foi gerida por responsáveis nomeados pelo Cliente X.

### 3.1.2 Organização do Trabalho em Equipa

O Cliente X. é composto por várias entidades constituídas por equipas de trabalho multidisciplinares consoante a sua área de trabalho, funcional ou técnica.

A gestão e coordenação do projeto foi assegurada pela gestão de projeto, integrando elementos responsáveis do Cliente X. e da Unipartner IT Services, aos quais responderam os restantes elementos da equipa de trabalho. Da parte da Unipartner IT Services, a equipa de trabalho é composta por um diretor de projeto, um gestor de projeto e dois consultores de BI. Por sua vez, da parte do Cliente X. tem-se, igualmente, um diretor e gestor do projeto e, ainda, a equipa de infraestruturas e a equipa de testes. A equipa encontra-se reunida no mesmo espaço físico, o que facilita a troca de impressões entre colegas, acelerando o processo de desenvolvimento. De forma a assegurar um entendimento comum do andamento do projeto por parte da Unipartner IT Services e do Cliente X., garantindo que eventuais desvios são prontamente identificados e tratados, preconiza-se a implementação de um efetivo plano de comunicação, estruturado em torno dos seguintes eixos:

- **Pontos de situação internos** que têm como finalidade reportar avanços e acontecimentos relevantes do projeto bem como os passos que ainda faltam realizar. É nestas reuniões, que a equipa foca objetivos e define metas para a realização do projeto.
- **Reuniões periódicas com o cliente**, onde são apresentadas e discutidas as implementações realizadas pela equipa de desenvolvimento. São nestas reuniões que o cliente pode alterar os requisitos do projeto bem como tomar decisões críticas na estrutura do produto e/ou outros pontos de interesse.

A comunicação entre as equipas foi realizada através da ferramenta Microsoft *Outlook* que diz respeito a um gestor de informação pessoal sendo parte integrante do Microsoft *Office*. Esta ferramenta permitiu o acesso ao correio eletrónico e aos calendários pessoal e da equipa. Foi com esta ferramenta que a calendarização dos projetos dentro da empresa foi realizada e onde todas as reuniões foram marcadas. Através de outra ferramenta da Microsoft, o Microsoft *Teams*, foi possível reunir toda a documentação necessária (apontamentos e anexos) num único ponto, acedido pelos dois elementos da equipa.

### 3.1.3 Processo de Desenvolvimento de *Software*

O desenvolvimento de *Software* foi orientado segundo o *Scrum*, uma Metodologia Ágil (Scrum Methodology, 2001). que segue uma abordagem iterativa e incremental para o desenvolvimento de *software* e gestão de projetos. Esta metodologia enquadrou-se bem com as características do projeto, cujos prazos de entrega foram curtos e os requisitos foram sofrendo alterações, ao longo do projeto. Tal como fora dito anteriormente, o projeto seguiu uma abordagem adaptativa e, por isso, as atividades que

constavam do plano não seguiram uma forma sequencial. A escolha desta metodologia de trabalho assentou no pressuposto de que os requisitos pudessem ser modificados pelo cliente a qualquer altura, sendo o projeto dividido em subconjuntos de funcionalidades, tal como aconteceu com o desenvolvimento dos processos de ETL e construção de relatórios. Enquanto a equipa aguardava validações por parte do cliente em relação aos *workshops* de apresentação dos primeiros relatórios (1ª *Release*) foram aperfeiçoados os processos de ETL, não seguindo a ordem normal de trabalhos.

A estrutura do *Scrum* adotada considerou a existência de três papéis, nomeadamente:

- **Product Owner**, que representa os interesses dos *stakeholders* e é o responsável pela definição de requisitos do produto, por atribuir prioridades do *product backlog*, por aceitar ou recusar o resultado de cada *sprint*. Este papel foi desempenhado pelo Cliente X.
- **Scrum Master**, responsável por assegurar a concretização dos objetivos definidos no planeamento, por facilitar a colaboração entre funções e áreas da equipa e por garantir que o processo de *Scrum* está a ser seguido. Este papel pertenceu ao gestor de projeto da parte da Unipartner IT Services e ao gestor de projeto do Cliente X.
- **Equipa de desenvolvimento**, grupo multifuncional de pessoas responsáveis, pelo desenvolvimento e entrega das funcionalidades do produto. Os elementos das equipas funcional e técnica da Unipartner IT Services representaram este papel.

Deste processo de desenvolvimento de *software* fizeram ainda parte dois tipos de artefactos:

- **Product Backlog**: é mantido pelo *product owner* e representa uma lista de requisitos solicitadas pelo cliente, cujos itens do *backlog* estão ordenados pela sua prioridade. Pode ser alterada pelo *product owner* a qualquer momento, desde que as alterações não afetem o período do *sprint* em curso.
- **Sprint Backlog**: uma sublista do *product backlog*, que contém os requisitos/funcionalidades que se pretende concretizar no *sprint* seguinte, nomeadamente os mais prioritários.

Resumindo, estes dois artefactos são diferentes, na medida em que o *product backlog* contém os requisitos de alto-nível do projeto e o *sprint backlog* informa sobre como a equipa do projeto irá implementar os requisitos nas diferentes alterações realizadas no projeto.

A organização do trabalho da equipa esteve bem alinhada com o processo de desenvolvimento de *Software* representado na Figura 8:





Figura 8: Processo de desenvolvimento de *Software*

Fonte: Adaptada de Figura disponível em: < <https://javafromdev.com/2017/09/15/scrum-methodology/> > Acesso em ago. 2018.

## 3.2 Identificação de Requisitos

O presente projeto teve como âmbito a criação de um *reporting* corporativo para o Cliente X., no qual se incluiu a integração de três fontes de dados operacionais e visou, numa primeira fase, conhecer as fontes de dados, as aplicações e infraestruturas que suportavam as fontes, como base para a definição do plano de criação de relatórios, em Microsoft *Power BI*. O diagnóstico dos requisitos do projeto teve como objetivo definir uma estratégia de análise de dados e arquitetura de implementação da solução, tendo por base um levantamento da situação atual da informação, estabelecer as dependências existentes entre as fontes de dados que se deveriam assegurar no decorrer do desenvolvimento do projeto (objetivo 1 – ver ponto 1.4.1 , página 8).

Numa primeira conversa com o Cliente X. ficou definida a avaliação da situação atual, identificação e caracterização das diferenças encontradas na análise dos dados e possíveis correções a realizar bem como a definição de prioridades de ação.

O Cliente X. dispõe de um número alargado de sistemas de informação, desenvolvidos para suportar as especificidades da sua atividade, que constituem as fontes de dados para o DW corporativo.

No âmbito da prova de conceito com os relatórios mensais referentes à fonte de dados A, foi exigida a construção de relatórios que respondessem a 9 análises, e que refletissem as necessidades do Cliente X.:

1. Número de Pedidos por Estado, com filtro por Tipo de Pedido, Cliente e Data (nesta análise foi necessário ter em conta a perspetiva histórica de 2 anos, para permitir comparações entre períodos homólogos e descer ao nível do mês);
2. Evolução Histórica do Número de Pedidos Abertos, com filtro por Tipo de Pedido, Cliente e Data;
3. Evolução Histórica do Número de Pedidos Fechados, com filtro por Tipo de Pedido, Cliente e Data;
4. Número de Pedidos em Carteira (à data da extração – posição atual), com filtro por Tipo de Pedido e Cliente (no relatório implementado, aplicou-se o filtro, primeiro por Estado e, depois por Cliente);

5. Antiguidade em Carteira, com filtro por Tipo de Pedido e Cliente (com possibilidade de descer ao nível do mês);
6. Número de Pedidos no Estado em Análise, com filtro por Tipo de Pedido e Cliente;
7. Número de Pedidos Abertos por Origem com filtro por Tipo de Pedido, Cliente e Data;
8. Número de Pedidos em Execução por Origem e Dimensionamento;
9. Número de Pedidos Fechados por Origem com filtro por Tipo de Pedido, Cliente e Data.

Quanto à segunda fonte de dados, o âmbito do relatório mensal incluiu sete relatórios, tendo em conta as seguintes variáveis: “organização”, “estrutura”, “área”, “tipo” e “estado do recurso” e do “projecto” e “tipologia de projecto”. A estas análises, foram implementados filtros relevantes: “tipo de projecto”, “mês”, “ano”, “tipologia de actividade”, “tipo de projecto”, entre outras:

1. Contabilização do número de projectos por estado;
2. Análise do número de projectos entregues por data;
3. Previsão do número de projectos entregues por data (com preferência para a apresentação do ano corrente mais o ano seguinte);
4. Distribuição de horas reportadas por tipo de actividade;
5. Distribuição de horas reportadas por tipologia de actividade;
6. Distribuição de horas reportadas por tipologia de recurso;
7. Distribuição de horas reportadas por tipologia de actividade (com um filtro por tipo de recurso).

Por último, a fonte de dados C albergou trinta e duas análises que constavam do âmbito do relatório mensal. A análise dos requisitos, modelação dimensional e construção dos relatórios em Microsoft *Power BI* referentes a esta fonte de dados ficou concluída no dia 31 de dezembro de 2017. Por ter sido, exclusivamente, da responsabilidade do colega de equipa da autora, não foi mencionado neste projeto.

Com o intuito de providenciar as análises requeridas, foram realizadas as seguintes tarefas: filtragem e verificação manual de informação desnecessária, de modo a obter dados corretos e inteligíveis, evitando, deste modo, informação de difícil compreensão; criação do diagrama de modelos dimensionais a serem implementados, após consulta das fontes de dados e estudo das relações existentes entre as mesmas; e implementação geral do modelo de dados. A autora, nesta fase, teve de perceber a relação entre os campos (colunas) e fazer o mapeamento com os relatórios (prova de conceito), disponibilizados pelo Cliente X., relativamente à fonte de dados A e, juntamente com o colega de equipa, perceber a relação entre os campos da fonte de dados B.

### 3.3 Caracterização dos dados do Processo (Modelação Dimensional)

Com base no levantamento supracitado, foi elaborada uma proposta de estrutura de modelos de dados que cobrissem, de uma forma abrangente e efetiva, as necessidades de negócio identificadas. Nesta fase do projeto, a autora ficou encarregue de analisar, identificar e sumarizar todos os requisitos necessários para a concretização do modelo de dados para a fonte de dados A. Algumas métricas foram implementadas e utilizadas nos relatórios em Microsoft *Power BI* para permitir que os utilizadores finais pudessem desenvolver outros relatórios a partir dos definidos.

Por forma a obter-se uma visão ampla e completa dos dados e das suas relações, bem como do contexto em que estes se encontravam inseridos, desenvolveu-se um modelo relacional com base nos campos definidos para os relatórios operacionais, no qual se detalharam as relações entre dimensões e possíveis tabelas de Factos. A realização deste modelo permitiu, também, reduzir o número de campos que seriam extraídos e que pudessem ser obtidos através de outras tabelas.

Foram realizados diversos pontos de situação nesta fase, com o intuito de validar a exatidão dos dados e perceber se estes estavam corretos; perceber se o modelo de análise definido suportava as análises do relatório mensal inseridas na prova de conceito e disponibilizada pelo Cliente X.; definir nomenclaturas dos campos na implementação dos processos de ETL; identificar outros requisitos, principalmente, outras análises e definir as regras de segurança e permissões de acesso aos relatórios. Este primeiro mapeamento entre as fontes e os requisitos teve como objetivo mostrar graficamente todos os campos que iriam ser utilizados na componente analítica.

Foi utilizado, pela autora, a ferramenta Microsoft Excel para a análise funcional da informação nos ficheiros das fontes de dados. Foram, então, produzidos ficheiros com informação estruturada e relevante para o projeto, sendo que esses ficheiros serviram, posteriormente, de base para a criação dos modelos analíticos. Esses ficheiros continham os campos escolhidos para cada fonte de dados, preenchidos com os respetivos dados, à data do dia anterior. Cada modelo analítico teve, numa versão final, uma tabela de Factos, que continha as métricas ou os valores passíveis de serem analisados e, um conjunto de dimensões específicas ou partilhadas, onde os diversos campos que as compunham tiveram origem em fontes de informação diferentes consoante a análise pretendida.

Para ser possível implementar as métricas e relatórios requisitados, foram criados dois diagramas em estrela (fonte de dados A e B), para cada um dos processos de negócio e integrados no DW. A apresentação das dimensões incluiu a identificação e descrição da dimensão, das suas propriedades e hierarquias envolvidas. O detalhe das métricas compreendeu a identificação, a descrição e a fórmula associada (quando aplicável).

Os dados do projeto corresponderam a três fontes de dados operacionais, tendo em conta a tipologia da atividade, com diferentes âmbitos de trabalho: fonte de dados A (atividade planeada), fonte de dados B (atividade planeada) e fonte de dados C (atividade corrente). Pelo explicado anteriormente, a fonte de dados C não será contemplada no documento (ponto 3.2 ).

### **3.3.1 Fonte de dados A**

A fonte de dados A (Aplicação de Gestão de Pedidos) é uma plataforma que cria e gere pedidos de desenvolvimento e de otimização, que suportam a atividade planeada do Cliente X., cujo documento de extração consiste em dois ficheiros Excel que contêm os dados existentes na fonte de dados, com a designação “Pedidos.csv” e “Tarefas.csv” e, com a extensão *Comma-Separated Values* (CSV). Na realização deste projeto, as análises recaíram apenas sobre o ficheiro “Pedidos.csv”, uma vez que este refletia todo o universo de dados que satisfaziam as necessidades supraditas no domínio do relatório mensal desenvolvido pelo Cliente X. O universo dos dados compreendeu o período entre 1 de janeiro de 2014 até ao momento. Este ficheiro .csv é constituído por 52 campos/colunas preenchidos com dados pertencentes à data da extração do dia anterior do vigente, de entre os quais foram utilizados para configurar as dimensões e tabelas de Factos existentes, bem como definir métricas de análises.

A partir do ficheiro .csv com os dados da fonte de dados A, foram definidos os campos mais importantes consoante as análises e requisitos do Cliente X., sendo que muitos deles constituíam métricas e análises sobre as dimensões. Numa primeira abordagem e de modo a replicar as análises do relatório mensal, ficou decidido que as colunas sobranes do ficheiro .csv, fossem colocadas na tabela de Factos. Mais tarde, na criação do modelo dimensional final, a autora teve de perceber a relação entre os campos e os relatórios até então criados e fazer o mapeamento para uma nova versão dos relatórios, criando, assim novas dimensões a partir dos campos analisados. Assim, definiram-se as primeiras dimensões e tabelas de Factos, de acordo com o explicado Figura 9:

Métricas	Sigla das Métricas	Dimensões	Sigla das Dimensões	Tipo Tabela de Factos	Descrição Tabela Factos	Periodicidade	Desagregações Disponíveis
Pedidos	M1	Data de Criação do Pedido	D1	Transaccional	Pedidos criados no período		Todas
Pedidos em Carteira	M2	Data de Fecho do Pedido	D2	Snapshot	Snapshot de pedidos	Diária	Cliente, Estado e Tipo de Pedido
		Tipo de Pedido	D3				
		Cliente	D4				
		Estado	D5				

	Métricas		Dimensões				
Análise	M1	M2	D1	D2	D3	D4	D5
Pedidos por estado com evolução histórica (2 anos)	x		x		x	x	x
Pedidos Fechados com evolução histórica (2 anos)		x	x		x	x	x
Pedidos Abertos com evolução histórica (2 anos)	x			x	x	x	x
Pedidos em Carteira	x		x		x	x	x
Antiguidade de Pedidos em Carteira		x	x		x	x	x
Pedidos em Estado de Análise	x		x		x	x	x
Pedidos em Estado de Execução	x			x	x	x	x
		x	x		x	x	x

Figura 9: Identificação Inicial dos Requisitos (Dimensões, Métricas e Tabelas de Factos) para a fonte de dados A  
Fonte: Elaborado pela autora

Através da Figura 9 e, numa primeira abordagem, foi possível verificar a existência de duas tabelas de Factos, em que uma era do tipo transaccional (“Pedidos”), onde se usaram métricas aditivas que podiam ser somadas em todas as dimensões, porque englobava uma análise dos dados e onde estavam presentes todas as dimensões e a outra era do tipo *accumulating snapshot* (“Pedidos em Carteira”), que se tratava de uma tabela de Factos que acompanhava um processo recorrente, de duração variável, neste caso o ciclo de vida de um pedido. Os dados novos, nestas tabelas, apareciam através de inserções e atualizações de linhas.

A ideia expressa na primeira abordagem que sugeria a existência de duas tabelas de Factos foi descartada, uma vez que uma tabela de Factos do tipo *accumulating snapshot* exigia diversas dimensões de data, onde cada uma delas representava o momento em que determinada fase do pedido fosse concluída. Ficou, então, decidido a criação de apenas uma, do tipo transaccional com uma granularidade tão fina quanto possível, com todas as dimensões relevantes e criada uma dimensão extra, “Data da Fotografia”, que guardasse o histórico e as análises diárias, permitindo a comparação entre períodos homólogos.

Assim sendo, o modelo de dados de BI (*Star-Schema*) apresentado na Figura 10 permitiu dar resposta às análises identificadas no âmbito bem como análises complementares, utilizando todos os campos de interesse provenientes da fonte de dados e validados pelo Cliente X. O modelo final de dados expõe a tabela de Factos (a verde) e dimensões de análise (a azul), sendo as análises possíveis representadas pelas setas entre estas entidades.

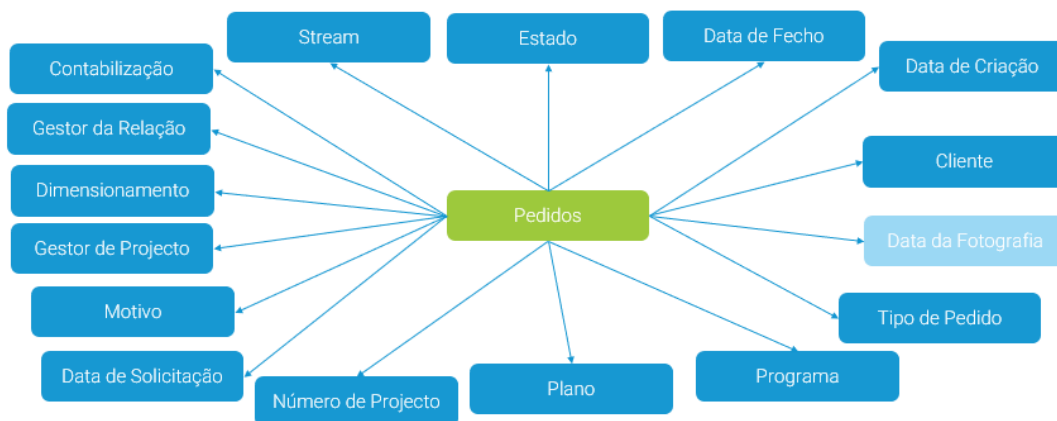


Figura 10: Modelo de Dados da Fonte de dados A  
Fonte: Elaborado pela autora

Este modelo é constituído por dezasseis dimensões, sendo estas: “Contabilização”, “Gestor da Relação”, “Dimensionamento”, “Gestor de Projecto”, “Motivo”, “Data de Solicitação”, “Número de Projecto”, “Plano”, “Programa”, “Tipo de Pedido”, “Data de Fotografia”, “Cliente”, “Data de Criação”, “Data de Fecho”, “Estado” e “Stream”. Atente-se que a dimensão “Data de Fotografia” não teve correspondência com os campos apresentados na fonte de dados, sendo apenas utilizada para a implementação de um mecanismo que permitiu o carregamento de dados históricos e, posterior, construção de relatórios em *Power BI*.

Todas as dimensões apresentadas são do Tipo 1 para o estudo das *Slowly Changing Dimension* (SCD) em que não se verifica alterações para armazenamento do histórico, ou seja, não há uma versão do registo modificado. Trata-se do tipo mais simples das dimensões, pois não há nenhum controlo específico para a atualização dos dados, havendo apenas sobreposição dos mesmos. Como exceção temos a dimensão “Estado” que apresentou a técnica do tipo 2, em que se tornou importante manter o histórico das mudanças de estado dos pedidos. Por forma a garantir o histórico desta dimensão, foram adicionadas três novas colunas na base de dados de *Enterprise Datawarehouse* (EDW) que permitiram guardar o histórico da informação dos estados, nomeadamente: “ValidFrom”, “ValidTo” e “IsActive”, em que cada linha tinha uma data de entrada, uma data limite e uma coluna “IsActive” que permitia aceder à linha mais atual do estado de um determinado pedido. Sempre que aparecia um registo novo, a informação era verificada com a existente na tabela de Factos da base de dados de EDW. Caso não existisse, era feita a inserção de um novo registo correspondendo à entrada de um novo dado. Caso existisse, a linha na tabela de EDW era colocada a “False” (=0) na coluna “IsActive” e a nova linha, era colocada a “True” (=1) com a data de entrada. Na camada de transformação do processo de ETL, o histórico associado à dimensão “Estado” é explicado mais detalhadamente. O histórico é apresentado na dimensão e no cruzamento de informação para a tabela de Factos.

Alguns exemplos de dimensões da fonte de dados A encontram-se referenciadas na Tabela 3 e Tabela 4, para melhor compreensão da apresentação das dimensões. As restantes encontram-se em apêndice (Apêndice II – Outras Dimensões existentes na Fonte de Dados A).

Tabela 3: Campos da Dimensão “Cliente”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKCustomer	<i>Integer</i>	PK	-	Chave primária (invisível)	“140”
AcronymCustomer	<i>Varchar (50)</i>	-	1	Entidade (visível)	“X Serviços”
CustomerDesignation	<i>Varchar (500)</i>	-	1	Estrutura (visível)	“Área de Engenharia de Aplicações”
Resource	<i>Varchar (500)</i>	-	1	Interlocutor (visível)	“Luís Figueiredo”

Fonte: Elaborado pela autora

A dimensão “Cliente” contém a informação referente ao cliente de um determinado pedido. Esta dimensão apresentou uma hierarquia com três níveis de detalhe:

“Entidade” -> “Estrutura” -> “Interlocutor”

Tabela 4: Campos da Dimensão “Estado”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKRequestStatus	<i>Integer</i>	PK	-	Chave primária (invisível)	“1”
RequestStatus	<i>Varchar (50)</i>	-	2	Estado (visível)	“Aberto”
RequestSub_Status	<i>Varchar (50)</i>	-	2	Sub-estado (visível)	“Em análise”
RequestSub_Sub_Status	<i>Varchar (50)</i>	-	2	Sub Sub-estado (visível)	“Registrado”
ValidFrom	<i>Date</i>	-	-	Este campo está invisível	“2018-04-16”
ValidTo	<i>Date</i>	-	-	Este campo está invisível	“9999-12-31”
IsActive	<i>Bit</i>	-	-	Este campo está invisível	“1”

Fonte: Elaborado pela autora

A dimensão “Estado” contém a informação referente ao estado em que se encontra um determinado pedido. Esta dimensão apresentou três níveis de detalhe:

“Estado” -> “Sub-Estado” -> “Sub Sub-Estado”

As dimensões referentes às datas (“Data de Fotografia”, “Data de Criação”, “Data de Fecho” e “Data de Solicitação”) apresentaram uma estrutura comum: Data; Ano; Ano /Semestre; Ano /Trimestre; Ano/ Mês; Ano/ Semana; Semestre; Trimestre; Mês; Número do Mês; Semana; Semana do Ano; Número do Dia da Semana; Dia; Dia da Semana; É dia da semana. Em todas as dimensões datas, estas apresentaram uma hierarquia, segunda a data, composta por três níveis:

Ano -> Mês -> Dia

Como as dimensões “Data” tinha uma estrutura semelhante, a solução aplicada passou pela criação de uma dimensão “Conformed.DimDate” na fonte de dados de EDW, mais tarde explicada, no ponto 3.5.1, transversal a todas as fontes de dados (Tabela 5):

Tabela 5: Campos da Dimensão “Data” (“Conformed.DimDate”)

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
DateKey	<i>Integer</i>	PK	Chave Primária	“20070101”
Date	<i>Date</i>	-	Data	“2007-01-01”
Day	<i>Integer</i>	-	Dia	“1”
Week	<i>Integer</i>	-	Semana	“1”
Month	<i>Integer</i>	-	Mês	“1”
Year	<i>Integer</i>	-	Ano	“2007”
Half	<i>Varchar (5)</i>	-	Semestre	“S1”
Quarter	<i>Varchar (5)</i>	-	Trimestre	“Q1”
YearHalf	<i>Varchar (10)</i>	-	Ano/Semestre	“2007/S1”
YearQuarter	<i>Varchar (10)</i>	-	Ano/Trimestre	“2007/Q1”
YearMonth	<i>Varchar (10)</i>	-	Ano/Mês	“2007/01”
YearWeek	<i>Varchar (10)</i>	-	Ano/Semana	“2007s01”
MonthName	<i>Varchar (30)</i>	-	Nome do Mês	“Janeiro”
WeekDay	<i>Varchar (30)</i>	-	Dia da Semana	“Segunda-feira”
isWeekDay	<i>Bit</i>	-	É Dia da Semana	“1”
NumWeekDay	<i>Integer</i>	-	Número Dia da Semana	“2”

Fonte: Elaborado pela autora

Definiu-se, nesta fonte de dados, uma única tabela de Factos, “Pedidos” (Tabela 6) e seis métricas disponíveis: “Número de Pedidos” (Contagem), “Quota Cativo” (soma), “Quota Efectivo” (soma), “Scoring” (valor médio), “Aberto” (contagem do número pedidos no estado “Aberto”) e “Fechado” (contagem do número de pedidos no estado “Fechado”):

Tabela 6: Campos da Tabela de Factos “Pedidos”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKRequest	<i>Integer</i>	PK	Chave Primária (invisível)	“1”
RequestNumber	<i>Varchar (15)</i>	-	Nº de Pedidos	“2014/005”
FKRequestType	<i>Integer</i>	FK	Este campo está invisível	“-1”
FKRequestStatus	<i>Integer</i>	FK	Este campo está invisível	“21”
FKCustomer	<i>Integer</i>	FK	Este campo está invisível	“919”
Creation Date	<i>Date</i>	-	Data de Criação	“2014-01-03”
Closing Date	<i>Date</i>	-	Data de Fecho	NULL
Request Date	<i>Date</i>	-	Data de Solicitação	NULL

SnapDate	<i>Date</i>	-	Data de Fotografia	“2014-05-30”
FKAccounting	<i>Integer</i>	FK	Este campo está invisível	“-1”
FKRelationshipManager	<i>Integer</i>	FK	Este campo está invisível	“-1”
FKSizingHour	<i>Integer</i>	FK	Este campo está invisível	“-1”
FKProjectManager	<i>Integer</i>	FK	Este campo está invisível	“-1”
FKReason	<i>Integer</i>	FK	Este campo está invisível	“-1”
FKProjectNumber	<i>Integer</i>	FK	Este campo está invisível	“-1”
FKPlan	<i>Integer</i>	FK	Este campo está invisível	“-1”
FKProgram	<i>Integer</i>	FK	Este campo está invisível	“-1”
FKStream	<i>Integer</i>	FK	Este campo está invisível	“-1”
CaptiveQuota (soma)	<i>Varchar (500)</i>	-	Quota Cativo	NULL
EffectiveQuota (soma)	<i>Varchar (500)</i>	-	Quota Efectivo	NULL
Scoring (valor médio)	<i>Varchar (500)</i>	-	Scoring	NULL
Aberto (contagem)	<i>Integer</i>	-	Abertos	“0”
Fechado (contagem)	<i>Integer</i>	-	Fechados	“1”

Fonte: Elaborado pela autora

A granularidade na tabela de Factos é o mais baixa possível, uma vez que quanto mais detalhados fossem os Factos, mais dimensões se conseguiam utilizar nas análises. Só assim se conseguia analisar os dados de todas as formas possíveis e retirar o máximo de informação acerca dos dados. A tabela de Factos é do tipo transacional e é a mais comum dos três tipos. Esta tabela registou uma linha por cada transação, ou seja, uma linha era criada sempre que ocorria um evento. Para a fonte de dados A, tornou-se necessário recriar o histórico relativamente às mudanças de estado dos pedidos, sendo utilizados dois conjuntos de ficheiros com formatos distintos:

- Formato A: Ficheiros equivalentes ao ficheiro “Fonte de dados A 20180109” e com dados compreendidos entre 1 de janeiro de 2017 até à data em que a solução fosse colocada em produção.
- Formato B: Ficheiros equivalentes aos ficheiros “Acc20160210” e “Acc20161228” e que compreendessem carregamentos anteriores de dados, de dia 1 de janeiro de 2014 a dia 1 de janeiro de 2017, exclusive.

Esta fonte de dados foi detalhadamente descrita e foram especificadas, em pormenor, as relações entre os campos, pois foi objeto de estudo da autora.

### 3.3.2 Fonte de dados B

A fonte de dados B é uma plataforma que permite o planeamento e a gestão de projetos, quando necessário, nas suas mais diversas vertentes, como planeamento de tarefas, alocação de recursos e reporte de horas. A análise desta fonte de dados ficou ao encargo do segundo elemento da equipa, sendo a sua análise acompanhada de forma indireta pela autora. Os dados extraídos da fonte de dados B são provenientes de uma base de dados e de um ficheiro em Microsoft Excel. As análises a realizar nesta



fonte, recaíram exclusivamente sobre os projetos que fossem do tipo “Projecto” (PR) e “Pequeno Projecto” (PP), tendo em conta a duração de realização dos projetos.

No modelo de dados, foram identificadas as métricas e respetivas agregações e os diversos eixos de análise (dimensões), tal como o implementado para a fonte de dados A.

Por forma a dar resposta às análises identificadas no âmbito, foi construído o modelo final de dados BI representado na Figura 11. O modelo expõe a tabela de Factos (a verde) e dimensões de análise (a azul), sendo que as análises possíveis estão representadas pelas setas entre estas entidades.



Figura 11: Modelo de Dados da Fonte de dados B

Fonte: Elaborado pela autora

Este modelo é constituído por dez dimensões, sendo estas: a dimensão “Projecto”, “Recurso”, “Cliente”, “Nível de Serviço”, “Data de Início de Projecto”, “Data de Fim de Projecto”, “Data Esperada de Fim de Projecto”, “Data de Reporte”, “Estrutura do Recurso” e “Tipo de Recurso”. Todas as dimensões apresentadas são do Tipo 1 para SCD em que não se verifica alterações para armazenamento do histórico, ou seja, não há uma versão do registo modificado.

Alguns exemplos de dimensões da fonte de dados B encontram-se apresentadas nas tabelas seguintes, para melhor compreensão da apresentação das dimensões:

Tabela 7: Campos da Dimensão “Cliente”

Campo	Tipo	Chave	SCD	Descrição do Atributo	Exemplo
PKClient	<i>Integer</i>	PK	-	Chave primária (invisível)	“2”
Client	<i>Varchar (50)</i>	-	1	Cliente (visível)	“X INFORMÁTICA» Y”
Client_Tier1	<i>Varchar (50)</i>	-	1	Cliente Nível 1 (visível)	“X INFORMÁTICA”
Client_Tier 2	<i>Varchar (50)</i>	-	1	Cliente Nível 2 (visível)	“Y”

Fonte: Elaborado pela autora

A dimensão “Cliente” contém a informação referente ao cliente de um determinado projeto. Esta dimensão apresentou uma hierarquia com dois níveis de detalhe:

“Cliente Nível 1” -> “Cliente Nível 2”

Tabela 8: Campos da Dimensão “Nível de Serviço”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKServiceLevel	<i>Integer</i>	PK	-	Chave primária (invisível)	“71”
Service	<i>Varchar (500)</i>	-	1	Serviço (visível)	“LS até 31-12-2014» NÃO UTILIZAR – Infra-Estrutura»IE0101- Sistemas”
Service_Tier1	<i>Varchar (500)</i>	-	1	Serviço Nível 1 (visível)	“LS até 31-12-2014”
Service_Tier2	<i>Varchar (500)</i>	-	1	Serviço Nível 2 (visível)	“NÃO UTILIZAR – Infra-Estrutura”
Service_Tier3	<i>Varchar (500)</i>	-	1	Serviço Nível 3 (visível)	“IE0101-Sistemas”

Fonte: Elaborado pela autora

A dimensão “Nível de Serviço” contém a informação referente ao nível de serviço de um determinado projeto. Esta dimensão apresentou uma hierarquia com três níveis de detalhe:

“Nível de Serviço 1” -> “Nível de Serviço 2” -> “Nível de Serviço 3”

Tabela 9: Campos da Dimensão “Tipo de Recurso”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKResourceType	<i>Integer</i>	PK	-	Chave primária (invisível)	“1”
ResourceType	<i>Varchar (255)</i>	-	1	Tipo de recurso (visível)	“Externo”
ResourceSubType	<i>Varchar (255)</i>	-	1	Subtipo de recurso (visível)	“Abs”

Fonte: Elaborado pela autora

A dimensão “Tipo de Recurso” contém a informação referente ao tipo de recurso exigido no desenvolvimento de projetos. Esta dimensão apresentou uma hierarquia com dois níveis de detalhe:

“Tipo de Recurso” -> “Sub-Tipo de Recurso”

Tabela 10: Campos da Dimensão “Estrutura do Recurso”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKResourceRBS	<i>Integer</i>	PK	-	Chave primária (invisível)	“1”
RBS	<i>Varchar (50)</i>	-	2	Estrutura de Recurso (visível)	“X» Y”

RBS_Tier1	Varchar (50)	-	2	Estrutura de Recurso Nível 1 (visível)	“X”
RBS_Tier2	Varchar (50)	-	2	Estrutura de Recurso Nível 2 (visível)	“Y”
RBS_Tier3	Varchar (50)	-	2	Estrutura de Recurso Nível 3 (visível)	Sem valor atribuído
RBS_Tier4	Varchar (50)	-	2	Estrutura de Recurso Nível 4 (visível)	Sem valor atribuído
RBS_Tier5	Varchar (50)	-	2	Estrutura de Recurso Nível 5 (visível)	Sem valor atribuído

Fonte: Elaborado pela autora

A dimensão “Estrutura do Recurso” contém a informação referente aos recursos usados em determinado projeto. Esta dimensão apresentou uma hierarquia com cinco níveis de detalhe:

“Estrutura do Recurso Nível 1” -> “Estrutura do Recurso Nível 2” -> “Estrutura do Recurso Nível 3” -> “Estrutura do Recurso Nível 4” -> “Estrutura do Recurso Nível 5”

Quanto às dimensões que envolviam datas (“Data de Fim de Projecto”, “Data de Início de Projecto”, “Data de Reporte” e “Data Esperada de Fim de Projecto”), estas apresentaram a estrutura da “Conformed.DimDate” apresentada no ponto anterior referente à fonte de dados A.

Nesta fonte de dados, definiu-se como tabela de Factos, as “Horas Reportadas”. A granularidade na tabela de Factos foi o mais baixa possível, tal como para a fonte de dados A. É uma tabela de Factos do tipo transacional, sendo este tipo, o mais comum dos três tipos existentes e é uma tabela que regista uma linha por cada transação, ou seja, sempre que ocorria um novo evento, era criada uma nova linha. As métricas disponíveis neste modelo de dados foram as seguintes: número de horas planeadas (soma), número de horas contabilizadas (soma), número de projetos (contagem) e número de recursos (contagem). Ainda, de acordo com o modelo de dados, é importante referenciar que as dimensões “Projecto” e “Recurso” se encontram detalhadas em apêndice (Apêndice III – Outras Dimensões existentes na Fonte de Dados B).

A informação referente à fonte de dados B foi reunida numa única tabela de Factos “Horas Reportadas” (Tabela 11):

Tabela 11: Campos da Tabela de Factos “Horas Reportadas”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKReportedHours	Integer	PK	Chave Primária (invisível)	“1”
FKResource (contagem)	Integer	FK	Este campo está invisível	“352”
FKReportedDate	Date	FK	Este campo está invisível	“2015-12-03”
ActualWorkBillable (soma)	Float	-	Horas Reportadas	“0.5”
PlannedWork (soma)	Float	-	Horas Planeadas	“0.5”
FKProject (contagem)	Integer	-	Este campo está invisível	“543”

Fonte: Elaborado pela autora

A autora acompanhou os trabalhos realizados em relação a esta fonte de dados, ficando responsável pela escrita da documentação referente a esta primeira análise dos dados, para validação por parte do Cliente X.

### 3.4 Arquitetura da Solução e seus principais Componentes

A solução desenvolvida esteve alinhada com o objetivo de facultar com rapidez, flexibilidade e de uma forma completa, toda a informação de gestão necessária à análise de negócio.

A arquitetura de solução consiste no desenho e na implementação de estratégias de tecnologia que visam suportar os objetivos de negócio da empresa, tendo em conta três sistemas independentes, com origens e características tecnológicas diferentes. **Neste projeto, a arquitetura foi definida segundo uma abordagem *on-premises*.** Uma solução de infraestrutura *on-premises* é uma solução que visa o uso de servidor e recursos de tecnologias de informação dentro da própria empresa ao invés do serviço remoto para processar as suas aplicações de *hardware* ou *software*. Compete, assim à empresa fazer configurações, customizações, implementações, atualizações, gerir o ambiente interno e cuidar da segurança e proteção dos dados corporativos. A escolha da arquitetura foi uma decisão dos gestores do projeto e da equipa de desenvolvimento em conjunto com o gabinete de segurança do Cliente X. Contudo, duas outras abordagens poderiam ter sido possíveis de implementar: abordagem na *Cloud* e uma abordagem híbrida (*Cloud + on-premises*). Não se optou por nenhuma destas duas abordagens, uma vez que iam contra a política de tratamento de dados da empresa, estipulado pelo gabinete de segurança da instituição. É importante referir, neste ponto, que a segurança de acesso à plataforma de BI foi concretizada com base em AD (*Active Directory*), integrada no servidor e que permitia aceder às contas de utilizadores. Assim o *Power BI* comunicou com os dados através de uma *gateway* explicada com mais detalhe a seguir.

A par destas decisões, foi realizado um documento técnico que providenciou a descrição da arquitetura da solução e que constituiu a base necessária para o planeamento, armazenamento e execução do projeto, tendo em conta as três fontes de dados existentes.

Nesta arquitetura, os dados provieram de três fontes de dados operacionais, foram trabalhados em três ambientes diferentes: desenvolvimento, qualidade e produção. O ambiente de desenvolvimento foi disponibilizado na fase de desenho técnico enquanto o de qualidade e produção foram disponibilizados durante a fase de implementação. As três fontes de dados existiam nos três ambientes, mas os dados não eram os mesmos. Assim, o ambiente de desenvolvimento foi o ambiente onde se criou as tabelas, *views* e *stored procedures*. O ambiente de qualidade serviu para testar tudo aquilo que foi desenvolvido, de modo a não comprometer os dados de produção. No ambiente de produção, os dados estavam prontos a serem usados, com todos os testes feitos e os *scripts* finais para a construção dos relatórios de apoio à decisão.

A arquitetura implementada no projeto assentou num conjunto de requisitos e pressupostos, os quais foram identificados nos pontos seguintes:

1. Existiu separação dos ambientes de qualidade e produção;
2. O desenvolvimento foi realizado no ambiente de desenvolvimento, sendo os dados passados depois para o ambiente de qualidade;

3. No ambiente de qualidade, existiu uma pequena amostra de dados iguais ao ambiente de produção, durante a fase de testes. Após os testes, os dados de qualidade foram removidos;
4. Os relatórios foram realizados tendo por base o ambiente de qualidade, os quais, posteriormente foram publicados para *Apps* de *Power BI* de qualidade. Quando os relatórios de qualidade foram aprovados, estes foram republicados através do *Power BI Desktop* com alteração do *DataSource* (uma configuração que permite a conexão entre as bases de dados e o servidor local da organização, onde se encontram as bases de dados com as fontes de dados), que referenciou o ambiente de produção, para *Apps* de *Power BI* de produção;
5. O consumo de relatórios foi realizado apenas através do *Power BI*;
6. O desenho dos relatórios foi feito sobre o *dataset* (conjunto de dados criados automaticamente no *Power BI*) existente, ou seja, os utilizadores com capacidade de produção de relatórios conseguiram comunicar via *Power BI Desktop* com o Serviço *Power BI*;
7. Os ambientes onde foi instalada a solução de BI comunicou com o serviço de *Power BI*;
8. Por fim, todos os utilizadores que necessitavam de ter acesso ao serviço de *Power BI* possuíam licenças de *Power BI Pro*.

A arquitetura implementada na solução final encontra-se esquematizada na Figura 12:

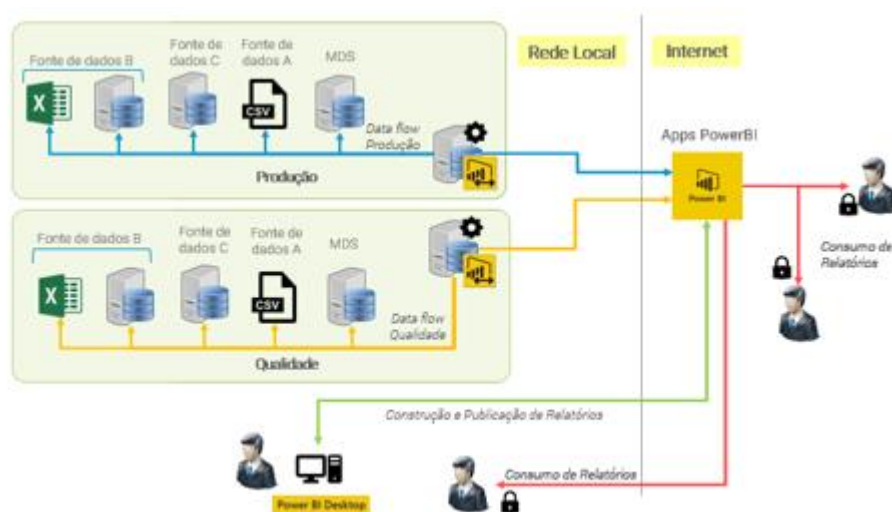


Figura 12: Arquitetura Física da Solução

Fonte: Elaborado pela autora

Na figura acima (Figura 12), foi possível verificar as diferentes iterações entre todos os componentes da solução, em que cada cor correspondeu a um bloco de solução, nomeadamente:

- *Data Flow* do ambiente de produção (Azul) e *Data Flow* do ambiente de qualidade (Laranja), responsáveis por realizar os processos de ETL e pela disponibilização dos dados das diferentes fontes sob a forma de relatórios;
- Construção e publicação de relatórios (Verde);
- Consumo de relatórios (Vermelho).

Em virtude do exposto, a Figura 13 representa a arquitetura de um *Data Flow* e identifica todos os passos desde a fonte de dados até a sua disponibilização via *Power BI*:

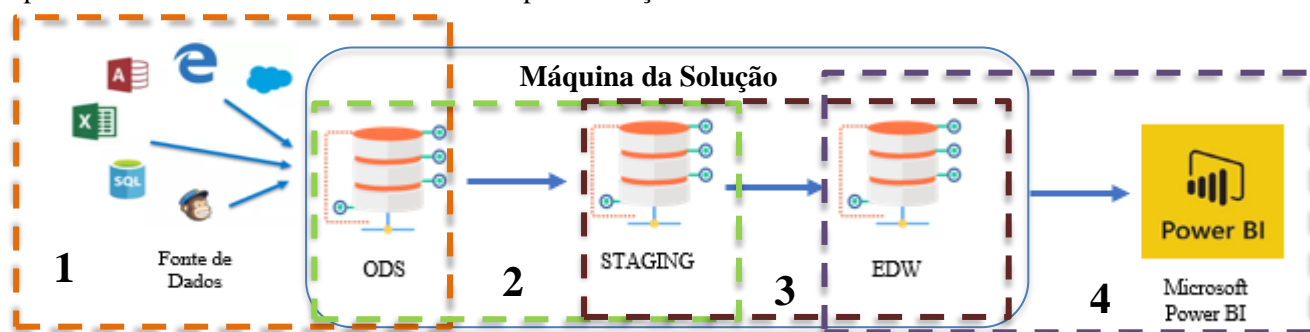


Figura 13: Arquitetura dos *Data Flows*

Fonte: Elaborado pela autora

Através da figura anterior, foi possível verificar que a arquitetura dos *data flows* era composta por 4 módulos importantes:

1. Extração – Camada de integração, responsável pela integração com as três fontes de informação que permitiu recolher os dados necessários para a realização das análises pretendidas. Para aceder a estas fontes de dados, foi necessário a definição de contas de acesso que permitiam que as máquinas locais/postos de trabalho do Cliente X. conseguissem aceder às fontes de dados, que estavam num *fileshare*, bem como a definição do local onde foram colocados/mapeados os diversos ficheiros que faziam parte das fontes e cujos campos precisavam de ser decodificados. Diariamente, os dados provenientes das fontes de dados operacionais foram extraídos diretamente para a base de dados *Operational Data Store* (ODS), pertencente à máquina da solução, responsável por armazenar uma cópia fiel dos dados que pertenciam às diferentes fontes de dados. Esta extração foi feita com recurso a pacotes de *Integration Services* que transferiram os dados da fonte de dados para a base de dados, sem que estes sofressem quaisquer transformações;
2. Transformação – As setas entre ODS e *Staging* implicaram uma camada de transformação. O processo de transformação consistiu em aplicar regras e funções sobre os dados extraídos, por forma a obter a estrutura pretendida no EDW. No contexto do projeto e já com os dados copiados para a base de dados ODS, foram aplicadas as regras de negócio. Continuamente, os dados foram transferidos para a base de dados de *Staging*, que constituiu a área de trabalho onde foram empregues as regras de negócio mais complexas e calculadas as métricas que suportavam as análises pretendidas;
3. Carregamento – Carregamento entre *Staging* e a base de dados de EDW: o processo de carregamento de dados correspondeu à última etapa do processo ETL e consistiu no carregamento das tabelas de destino, que estavam organizadas de acordo com o modelo de DW escolhido. Este processo foi realizado aquando da finalização do tratamento dos dados das tabelas de dimensões e após a criação das tabelas de Factos para serem inseridas

na base de dados relacional. O carregamento desta solução de BI ocorreu da base de dados de *Staging*, onde os dados foram transformados e limpos, para uma base de dados de EDW, onde ocorreu o seu armazenamento;

4. Carregamento para o Serviço Power BI – A última etapa destes *Data Flows* consistiu no carregamento dos dados da base de dados de EDW para o serviço *Power BI*, de forma diária, para que fosse possível o consumo de relatórios com a informação referente ao dia anterior. Este carregamento foi conseguido utilizando a *on-premises data gateway* instalada na máquina da solução, e que permitiu a comunicação entre a máquina e o serviço de *Power BI*. O carregamento dos dados foi feito por *Import* e não por *DirectQuery* (ligação direta às bases de dados com uma visão atual dos dados), uma vez que o projeto englobava um conjunto de dados pequenos. Outros motivos prenderam-se com o tratamento das colunas calculadas (hierarquia de datas, por exemplo) e o cálculo de métricas, recorrendo a DAX (consultas muito lentas), não serem suportados no modo *DirectQuery* (Microsoft Docs, DirectQuery, 2018). No carregamento dos dados para o serviço de *Power BI*, foi possível configurar a atualização agendada dos dados, para, por exemplo, voltar a importá-los todos os dias. Os relatórios foram atualizados automaticamente, recorrendo a um *job* sempre que o conjunto de dados fossem atualizados. A atualização dos dados no serviço de *Power BI* foi feita com base no esquema presente na figura, sendo os seus passos enumerados de seguida (Figura 14):

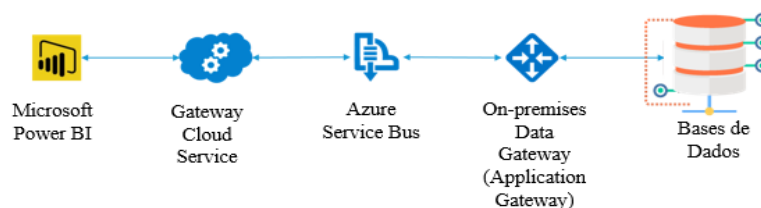


Figura 14: Atualização dos dados no serviço *Power BI*

Fonte: Elaborado pela autora

1. No serviço de *Power BI*, foi configurada a atualização dos dados através de um *Job*, definindo a periodicidade pretendida, a *gateway* (é um *software* que se instala dentro de uma rede local e que facilita o acesso aos dados nessa rede) a utilizar e qual a fonte de dados escolhida.
2. Quando a atualização dos dados foi desencadeada, o serviço *Power BI* criou uma mensagem e enviou-a para uma fila de espera.
3. O *gateway Cloud Service* é um serviço que contém o registo da *on-premises data gateway*, e foi responsável por analisar a mensagem e enviá-la para o *Azure Service Bus*.
4. A *on-premises data gateway* contactou o *Azure Service Bus* em busca de mensagens pendentes e, quando estas existissem, a mensagem era analisada e a *query* era executada na fonte de dados.
5. A fonte de dados, por sua vez, respondeu à *on-premises data gateway* com o resultado da *query* e, no preciso momento, o percurso inverso foi iniciado até que o resultado chegasse ao serviço de *Power BI*.

Relativamente ao *Job* criado no serviço de *Power BI*, este foi criado com o intuito de atualizar os dados, uma vez que a ligação do *Power BI* aos dados aconteceu através de *Import Data*, tal como explicado anteriormente. Tendo em conta este facto, foi necessário ter em consideração dois pontos:

- O serviço de *Power BI* pode levar até 15 (quinze) minutos até iniciar o processo de atualização dos dados.
- Após dois meses de inatividade, este *Job* será colocado em suspenso, até nova ativação. Um *dataset* é considerado como inativo, quando neste período, nenhum *dashboard* ou relatório construído sobre o mesmo é visitado.

Esta arquitetura apresentou uma fiabilidade comprovada e permitiu adicionar funcionalidades e acompanhar as necessidades resultantes da evolução dos sistemas operacionais paralelos. Neste ponto do projeto, a autora participou juntamente com o parceiro de equipa e gestor do projeto em diversas reuniões para decidir a arquitetura a implementar e quais as diretrizes a seguir.

## 3.5 Integração

Antes do desenvolvimento da camada ETL, foram realizados, numa primeira fase, *workshops* funcionais de modo a disponibilizar as análises do relatório mensal em *Power BI* bem como novas análises identificadas. A par da realização destes *workshops*, foi realizado o desenvolvimento da camada ETL, do repositório de DW e a disponibilização de novas versões dos relatórios mensais com *feedback* recebido pelo Cliente X.

Neste capítulo, é apresentado o desenvolvimento dos processos de ETL e modelação do repositório de DW do sistema de BI implementado na organização. Inicia-se o capítulo com a descrição do trabalho desenvolvido nos processos de ETL, por forma a transformar e a integrar os dados. Posteriormente, é descrita a conceção e implementação do DW que vai servir de suporte ao processo de tomada de decisão.

### 3.5.1 Processo de ETL

Da realização de várias reuniões com o cliente, foi possível conhecer melhor os dados e a forma de realização dos relatórios e especificar as necessidades a suportar pelo DW a construir. Posteriormente, foi possível gerar uma base de dados (ODS) a partir do modelo físico da arquitetura da base de dados fornecida pelo Cliente X. Neste momento, conhecida a estrutura da base de dados, as suas tabelas e os seus campos, definiram-se os elementos que vão sustentar a DW. Para tal, foi criado um *script* de dados para a inserção na base de dados dos ficheiros e *queries* pertencentes às bases de dados. Identificados e reunidos os elementos importantes para cada fonte de dados, procedeu-se à identificação das estrelas, dimensões e tabelas de Factos. A partir do conhecimento dos dados, procedeu-se à criação do modelo concetual de DW. Os diagramas em estrela que refletem as duas fontes de dados (A e B) foram descritos no ponto anterior.

A ferramenta utilizada para concretizar o processo de ETL foi o SSIS (objetivos 2 e 3 – ver 1.4.1, página 8), que permitiu a construção de fluxos de dados e tarefas de extração, transformação, integração e carregamento dos dados.



O primeiro passo foi desenvolver os processos de ETL das tabelas de dimensão e, posteriormente, os processos de ETL para as tabelas de Factos. Relembra-se que cada uma das tabelas de Factos está relacionada com um conjunto de tabelas de dimensões através de chaves estrangeiras.

O processo de ETL iniciou-se com a extração dos dados.

No SSMS, o caminho de elaboração das bases de dados teve em conta a seguinte metodologia (Figura 15):

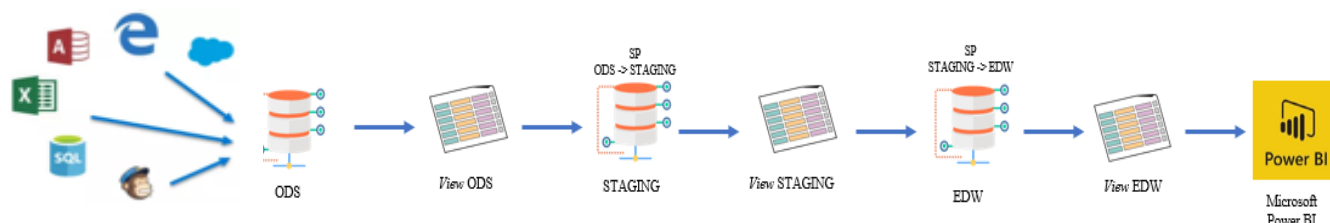


Figura 15: Metodologia do Processo de ETL  
Fonte: Elaborado pela autora

### 3.5.1.1 Extração

A etapa da **extração** corresponde ao processo de obtenção de uma cópia dos dados que estão na fonte em diversos formatos e estruturas. Em relação à etapa da extração será referenciado, com maior detalhe o caso da fonte de dados A, por ser da responsabilidade da autora e, com menor detalhe, as restantes fontes de dados.

Diariamente, os dados provenientes das fontes de dados operacionais correspondentes são extraídos diretamente para a base de dados ODS, pertencente à máquina da solução. Numa primeira abordagem, os ficheiros pertencentes às fontes de dados foram carregados através da criação de um projeto de *Integration Services*, com ligação à base de dados do projeto através do comando *Connection Manager* (*flat file source* para o ficheiro csv da tabela de Factos “Pedidos” da fonte de dados A) e comando *OleDb Destination* para uma base de dados criada temporariamente, de modo a ter visível os dados para a criação dos primeiros relatórios temporários em Microsoft *Power BI*, para validação com o cliente. Nesta fase, o processo de carregamento de dados provenientes da fonte de dados não foi automático, uma vez que não estavam definidas as contas de acesso às bases de dados local e aos ficheiros de Excel. Este facto, implicava que cada vez que existisse uma alteração dos ficheiros que continha a informação dos dados, era necessário carregar novamente os ficheiros, por intermédio de um *backup* à tabela “Requests” (tabela de Factos: “Pedidos”) construída no SQL Server. Mais tarde, foram cedidas as contas de acesso às bases de dados das fontes de dados.

O processo de extração da fonte de dados A ocorreu diariamente, em que o ficheiro “Pedidos.csv” foi extraído contendo a fotografia correspondente à data de extração. Posteriormente, este ficheiro foi carregado para a base de dados ODS que constituiu as bases das análises pretendidas para a fonte de dados A. Além disso, foi requerido pelo Cliente X., a recriação do histórico das mudanças de estado, as quais foram carregadas, uma única vez, no início, contendo os ficheiros, a posição de fim de mês desde 2014 até à data atual, permitindo obter o histórico do número dos pedidos em carteira no final de cada mês. Na fonte de dados A, foram inseridos os dados históricos de 2014 a 2017 através de uma funcionalidade do SQL Server, o “*Bulk Insert*” que é um código (Código 1) utilizado para importar e exportar dados a partir de bases de dados desenvolvidas em Microsoft *SQL Server Database*.

Por forma a integrar o histórico foi feito um mapeamento entre as informações passadas e os dados no momento, como se pode ver de seguida na etapa da transformação de dados. Tal como dito anteriormente, os ficheiros de histórico foram carregados, uma única vez, recorrendo à funcionalidade “*Bulk Insert*”, através de uma SP designada “uspCarregamentoHistorico” para a fonte de dados A:

--Excerto de Código para o Carregamento de Ficheiros de Histórico--

```

1 USE [ODS]
2 GO
3
4 /***** Object: StoredProcedure [Fonte de dados A].[uspCarregamentoHistorico]  Script Date: 01/09/2018   5 7:23:13 *****/
5
6 SET ANSI_NULLS ON
7 GO
8
9 SET QUOTED_IDENTIFIER ON
10 GO
11
12 -- =====
13 -- Author:                <Patrícia Martins>
14 -- Create date: <2018/02/01>
15 -- Description:    <Description,>
16 -- =====
17 CREATE PROCEDURE [Fonte de dados A].[uspCarregamentoHistorico]
18     -- Add the parameters for the stored procedure here
19
20 AS
21 BEGIN
22     -- SET NOCOUNT ON added to prevent extra result sets from
23     -- interfering with SELECT statements.
24     SET NOCOUNT ON;
25
26 --TRUNCATE TABLE --
27 Truncate Table [Fonte de dados A].[Requests_14-16]
28 Truncate Table [Fonte de dados A].[Requests_17-18]
29 --IMPORTS A DATA FILE INTO A DATABASE: Requests_14-16 e Requests_17-18--
30 --Ficheiro 2014.xlsx - 30/05/2014--
31 Bulk Insert [Fonte de dados A].[Requests_14-16]
32 From 'C:\Users\c89599394\Desktop\2014 csv\2014.csv' --existe um ficheiro para cada posição do mês--
33 With
34 (
35 FIRSTROW=4,
36 FIELDTERMINATOR ='\t',
37 ROWTERMINATOR = '\n',
38 CODEPAGE = 'ACP'--para caracteres especiais--
39 )

```

Código 1: Excerto de Código para o Carregamento de Ficheiros de Histórico da Fonte de Dados A  
Fonte: Elaborado pela autora

As linhas 27 e 28 permitiram fazer uma limpeza dos dados existentes nas duas tabelas que guardavam o histórico, sendo que depois foram inseridos os ficheiros do Cliente X. referentes ao histórico desde o dia 1 de janeiro de 2014. O exemplo da linha 31 diz respeito aos ficheiros carregados para a tabela em relação aos anos de 2014 a 2016, por serem ficheiros com a mesma estrutura de campos. Da linha 33 à 38, foram refletidas as condições necessárias para a passagem dos ficheiros, nomeadamente: tirando os cabeçalhos dos ficheiros, os dados começaram na linha 4 do ficheiro (representado na linha 35 do código). Ficou definido que o elemento de separação de colunas seria o

símbolo “£” (linha 36) por o símbolo “;” ser já um elemento de separação de informação dentro de alguns campos. Por fim (linha 38), houve a necessidade de implementar o código de página “ACP” que é o definido para caracteres especiais, admitindo as palavras acentuadas.

Diariamente, a informação da fonte de dados B, é proveniente de uma base de dados, com base nos pacotes de *Integration Services* fornecidos pela gestão de serviço, os quais contêm as *queries* de extração já definidas e a partir de ficheiros Excel, sendo a informação anterior substituída. Nessa perspetiva, o *reporting* terá sempre a posição do dia anterior.

O processo para extrair os dados, das tabelas que deram origem à tabela de Factos, ocorreu de igual modo ao das tabelas de dimensão e ocorreu na base de dados de ODS, responsável por ir buscar a estrutura dos campos que pertencem à fonte de dados, sendo que, posteriormente, foram construídas *views* sobre estas tabelas, permitindo construir as dimensões e tabelas de Factos. Por último, nesta base de dados figuraram ainda *Stored Procedures* que permitiram a funcionalidade “*Truncate*” às várias tabelas criadas nesta base de dados, por forma a limpar os dados inseridos e mantendo a estrutura das tabelas. A construção das *queries* em *views* de extração consistiu, ainda, no uso da instrução “*Distinct*” para retornar apenas valores distintos, que não se repetiam, visto que nas dimensões, não é possível ter dados duplicados.

Concluída a etapa de extração, seguiu-se a etapa de transformação em que os dados foram canalizados na DSA para percorrer todo o fluxo de transformação. Aqui os dados, sofreram as transformações necessárias e foram integrados na base de dados de *Staging*.

### 3.5.1.2 Transformação

A etapa da **transformação** consistiu em converter os dados em informação, através da manipulação dos formatos de dados e sua limpeza. Foram aplicadas determinadas regras e funções aos dados extraídos, tendo em vista a sua adequação e obtendo a estrutura pretendida na base de dados de EDW. Algumas transformações foram concretizadas através de *views* criadas na base de dados de ODS. Os restantes processos de transformação e correção de dados ocorreram na base de dados de *Staging*. Na base de dados de *Staging*, foram criadas as mesmas tabelas que na base de dados de ODS, pois foi nesta fase que ocorreram as transformações e correções de dados.

Nas *views* de *Staging*, foram criadas *views* para as tabelas de factos de modo a interligar as informações provenientes do ODS. As *Stored Procedures* criadas na base de dados de *Staging* foram utilizadas para passar os dados da base de dados de ODS para a base de dados de *Staging*. Depois de concluído todo o processo de transformação, os dados que se encontravam na DSA foram carregados para o DW, através do mapeamento entre os sistemas e respeitando as questões de integridade entre ambos.

Quanto às dimensões, estas tal como fora dito na revisão da literatura e na construção dos modelos de dados, apresentaram dois tipos para a atualização de dados já existentes. No tipo 1, ocorreu apenas uma atualização dos campos sem interessar o histórico. Por sua vez, a dimensão “Estado” exigiu um tratamento do tipo 2, em que o histórico foi adicionado e a informação antiga mantida. Para as dimensões deste último tipo, foram acrescentadas três colunas à tabela em EDW: “*ValidFrom*”, “*ValidTo*” e “*IsActive*”. Quando existiu inserção de uma nova linha, a primeira coluna foi preenchida com a data de inserção dos dados, sendo que a segunda coluna foi preenchida com a data “9999-12-31” não existindo uma validade definida. Por último, a última coluna assumiu o valor “*True*” que é igual a “1”, por se referir à linha com a informação mais atualizada. Estas três colunas diferiram o histórico da informação atual, uma vez que quando “*IsActive*” se encontrava a “*False*” (=0) representava uma linha contendo a informação histórica, pelo contrário quando esta coluna estava a “*True*” (=1) assumia-se

como a linha mais atual. Também pelas datas, foi possível identificar o período em que determinada linha esteve ou não ativa.

Esta inserção e atualização de linhas de histórico, principalmente referente às mudanças de estado dos pedidos da fonte de dados A, foi possível na tabela de factos, através da funcionalidade “Merge” que permitiu num único comando fazer o *Insert* e o *Update* dos dados. Como tabela fonte tivemos a tabela em *Staging* e, como tabela alvo, a de EDW, que constituiu a tabela final com os dados transformados, limpos e atualizados.

Verificou-se, ainda, que a estrutura dos ficheiros históricos da fonte de dados A era diferente entre eles e entre o modelo construído sobre a fonte de dados base. Após a análise dos campos existentes no ficheiro “Fonte de dados A 20180109”, a lógica a aplicar sobre este tipo de ficheiros e as respetivas colunas, consistiu na remoção de todos os caracteres iniciais até ao “;#”, inclusive. Depois de realizada esta limpeza, os dados foram carregados e, no caso de existirem novos valores nestes ficheiros que não constassem nas dimensões atuais, estes foram considerados como novos membros. Relativamente ao ficheiro “Acc20160210” e “Acc20161228” e aos restantes do mesmo tipo, o conteúdo de algumas colunas não correspondia diretamente ao conteúdo existente nas dimensões atuais.

Em relação ao campo “Estado”, tornou-se imprescindível o mapeamento entre os dados que compunham este campo datados antes de 1 de janeiro de 2017 e os dados posteriores a essa data, por forma a satisfazer os critérios exigidos para a primeira análise dos relatórios mensais. Ainda em relação ao “Estado”, não se verificou correlação entre alguns elementos dos ficheiros de histórico e alguns definidos no modelo de dados atualmente implementado. Posto isto, para os campos cujo mapeamento direto não foi possível, ficou definido uma estrutura com a hierarquia delineada anteriormente (Estado -> Sub-Estado -> Sub Sub-Estado), de modo a existir correlação entre os dados históricos e o modelo de dados.

No carregamento dos dados históricos referentes aos campos “Tipo de Pedido”, “Entidade” e “Estrutura”, não era obrigatório existir um mapeamento direto, pois os valores para estes campos já sofreram, ao longo dos anos, várias mudanças. Contudo, para estes campos, quando se verificou a existência de um mapeamento direto seguiu-se a abordagem do caso 1. Para os casos em que não existia mapeamento definido, esse elemento foi considerado como um novo valor da respetiva dimensão (caso 2).

### **Caso 1:**

Exemplo: Neste exemplo (Tabela 12), foi possível verificar que existia um mapeamento direto e bem definido entre os ficheiros “Acc20160210” e “Acc20161228” e o modelo de dados atualmente implementado. Nesta linha, o campo antes intitulado de “Solicitado por” tinha correspondência com o campo “Estrutura” do modelo de dados implementado assim como os dados que o constituíam. Ainda no mapeamento da dimensão “Cliente” pertencente à fonte de dados A, foi adicionada uma linha para cada “Entidade” e “Estrutura” com o “Recurso” igual a “Sem Recurso (Histórico)” quando não existia mapeamento entre os dados históricos e atuais.

Tabela 12: Mapeamento Direto entre os Ficheiros de Histórico e o Modelo de Dados implementado

	<b>Ficheiro Acc20160210 (Formato B)</b>	<b>Modelo de dados implementado</b>
<b>Campos</b>	<b>Solicitado por</b>	<b>Estrutura</b>
<b>Dados</b>	X de ALCÁCER-SAL E MONTEMOR-NOVO C.R.L.	6020 - Alcácer do Sal e Montemor-o-Novo

Fonte: Elaborado pela autora

### **Caso 2:**

Para todas as colunas que não fosse fornecido um mapeamento adequado, os dados históricos foram carregados e todos os dados não mapeados foram considerados como novos elementos das respectivas dimensões. Exemplo: Na dimensão “Tipo de Pedido”, na falta de mapeamento direto entre as opções para o “Tipo de Pedido”, a solução seguida consistiu em acrescentar uma nova linha a esta dimensão com o novo valor para o elemento inexistente (Tabela 13):

Tabela 13: Mapeamento entre os Ficheiros de Histórico e o Modelo de Dados implementado para a Dimensão “Tipo de Pedido”

Modelo de dados implementado		Modelo de dados implementado
Tipo de Pedido		Tipo de Pedido
Manutenção Evolutiva / Nova Aplicação		Manutenção Evolutiva / Nova Aplicação
Manutenção Correctiva - Reactiva		Manutenção Correctiva - Reactiva
Manutenção Correctiva - Pró-Activa		Manutenção Correctiva - Pró-Activa
Informação a Terceiros		Informação a Terceiros
Sistemas e Comunicações		Sistemas e Comunicações
Via Verde		Via Verde
Organizacional		Organizacional
Seleção de <i>Software</i>		Seleção de <i>Software</i>
		Caso Agregado

Fonte: Elaborado pela autora

Assim sendo, para todas as colunas em que não fosse fornecido um mapeamento adequado, os dados históricos foram carregados e todos os dados não mapeados foram considerados como novos elementos das respectivas dimensões.

Em EDW, não ocorreu transformação de dados, sendo que estes ficaram prontos para serem acedidos por um utilizador final após serem carregados para o Microsoft *Power BI*. Antes, foram criadas *views* com a informação que se pretendia mostrar ao utilizador final. Também nesta base de dados, os SP usados permitiram a passagem da base de dados de *Staging* para EDW.

### **3.5.1.3 Carregamento**

A etapa do **carregamento** dos dados da DSA para o DW correspondeu à última etapa do processo de ETL e foi realizada após a finalização do tratamento dos dados. O carregamento desta solução de BI ocorreu da base de dados de *Staging* para uma base de dados de EDW em que ocorreu o armazenamento de dados. Seguidamente, os dados foram carregados de EDW para o serviço de *Power BI*. Esta comunicação foi realizada na mesma máquina e instância de SQL, pelo que não foi necessária nenhuma comunicação via *Internet*.

A instalação em qualidade e produção incluiu o desenvolvimento de pacotes de instalação definidos pela equipa de testes do Cliente X. Na instalação em qualidade, foi necessário definir um conjunto de regras e princípios: fonte de dados A (localização e contas de acesso), fonte de dados B (definição das contas de acesso às bases de dados e definição das contas de acesso e local ao Excel) e fonte de dados C (definição das contas de acesso às bases de dados).

A autora neste ponto ficou encarregue das seguintes funções:

- Identificação das tabelas que correspondiam à extração, nomeadamente as tabelas de mapeamento e quais eram as tabelas auxiliares, para serem carregadas através do MDS, para que alguns dos seus campos fossem decodificados. As tabelas de mapeamento carregadas para o MDS encontram-se em apêndice (Apêndice IV – Tabelas de Mapeamento para o MDS);
- Continuação do desenvolvimento da base de dados ODS, *Staging* e EDW (focado na fonte de dados A);
- Criação de tabelas, vistas, procedimentos, montando os processos necessários para chegar desde o ficheiro até ao último modelo definido.

Concluído todo processo de ETL, os dados foram carregados no modelo de DW. Por fim, criou-se um *Job*, no *SQL Server Agent*, responsável pela execução do processo de ETL, nomeadamente dos *packages* implementados. Este *job* apresentou uma execução com uma periodicidade diária definida, neste caso durante a noite, no servidor onde se encontrava armazenado o DW, não condicionando desta forma a *performance* dos sistemas. O *job* criado permitiu a junção de todo o processo até ao momento construído, de modo a executar em determinado tempo todas as funcionalidades pretendidas.

Foram realizados pacotes de instalação de acordo com as indicações do cliente para o ambiente de qualidade que é semelhante ao de produção, onde foram realizados determinados testes. Posteriormente, foram definidos novos pacotes de instalação para o ambiente de produção, terminando a parte de manuseamento dos dados.

No final dos trabalhos de acesso, transformação e carregamento dos dados, foi criado um *Master Package* com as diretivas do processo a correr, nomeadamente os pacotes a executar, a sua ordem e os demais parâmetros associados. A criação deste *Master Package* foi necessária porque tradicionalmente, o processo de ETL foi desenvolvido por blocos de instruções, independentes entre si, num mesmo período de tempo e sem qualquer estrutura lógica de processos. Assim, os dados foram tratados e carregados consoante as necessidades identificadas.

### 3.5.2 Datawarehouse

Depois de realizada a caracterização dos dados operacionais e das respetivas fontes de dados, a modelação do DW é apresentada nesta seção (objetivos 4 e 5).

Neste projeto foi adotada a arquitetura de Ralph Kimball, em que os *Data Marts* se relacionam entre si. Para ser possível relacionar esses *Data Marts*, as dimensões comuns devem ser conformes e devem suportar a capacidade de integrar e analisar dados provenientes de múltiplos processos de negócio. Para serem consideradas conformes, assegurou-se que, nas tabelas de dimensão, as chaves,

nomes das colunas, definições dos campos selecionados e os seus valores eram consistentes em todos os processos de negócio.

A modelação multidimensional é considerada uma das etapas mais importantes e é um dos fatores críticos de sucesso num sistema de DW. Neste sentido, o modelo de dados proposto assentou na construção de uma estrutura que suportasse os principais indicadores de negócio, partindo dos requisitos e dos objetivos pretendidos para as análises do negócio. Ao todo, o modelo de DW integrou quatro tabelas de Factos e sessenta e duas dimensões, contando com as pertencentes à fonte de dados C. As tabelas de Factos foram representadas por: “Pedidos” (fonte de dados A); “Horas Reportadas” (fonte de dados B); “Incidentes” e “Pedidos” (fonte de dados C).

É importante referir que por questões de desempenho não foram utilizadas chaves compostas, optando-se por utilizar uma chave substituta (*Surrogate Key*) definida como chave primária, para cada uma das dimensões, baseada num número inteiro que foi atribuído sequencialmente (0, 1, 2, 3, ...) à medida que novos registos foram sendo inseridos. As chaves substitutas foram utilizadas para relacionar as tabelas de Factos com as tabelas de dimensão. Os elementos “ID” de cada dimensão permitiram associar todos os registos guardados na DW e obter informação mais detalhada quando necessário.

### 3.5.3 Avaliação de Segurança dos Dados

Relativamente às regras de segurança de dados, ficou definido que não existia nenhum tipo de segregação de dados tendo em conta o perfil de cada utilizador, ou seja, não era necessário a implementação de *Row-Level Security* (RLS) nem uma definição de grupos de acesso aos relatórios e *Workspaces*, a nível do Microsoft *Power BI*, sendo criado um único *Workspace* acedido livremente por qualquer colaborador. Por forma a que as fontes de dados fossem acessíveis, foi necessária a definição das contas de serviço a serem utilizadas, bem como o local onde foram colocados/mapeados os diversos ficheiros. Após algumas reuniões com o Cliente X, determinou-se, através de uma listagem dentro da própria organização, que alguns interlocutores ficariam responsáveis por criar e editar os relatórios apresentados e outros apenas os consultariam. Deste modo, os modelos de dados foram construídos com base no modelo de funções e responsabilidades já existentes na organização.

Por forma a seguir regras de segurança exigidas no tratamento sistemáticos de dados pessoais propostos pelo *General Data Protection Regulation* (GDPR), com início efetivo a 25 de maio de 2018, neste projeto, ficou garantido que sempre que acontecessem transferências ou cópias, estas tinham de ocorrer de forma isolada, por forma a assegurar a disseminação dos dados. Também os *backups* deveriam ser igualmente encriptados caso apresentassem dados pessoais, assim como o canal de comunicação utilizado. Ficou ainda estipulado que os ambientes não produtivos (desenvolvimento e qualidade) não deveriam utilizar dados pessoais. Em reuniões com o departamento de segurança do Cliente X, ficou clarificado que os dados não eram considerados sensíveis, que a sua transferência ocorria de forma segura e eram do conhecimento da comunidade de trabalho do cliente. Por esta razão, não foi necessário implementar o controlo de anonimização dos campos, pseudonimização dos campos, encriptação do repositório nem foi necessário mascarar os dados na apresentação para todos os ambientes.

Por último, o serviço de *Power BI Desktop* não apresentou acesso ao ambiente de produção, como defendido nos pressupostos da arquitetura da solução.

Assim sendo, a solução implementada disponibilizou um modelo de segurança flexível.

### 3.6 Camada de Visualização: Relatórios e *Dashboards* de BI

Para atingir esta meta, foi necessário efetuar grandes transformações dos dados provenientes dos sistemas operacionais, de forma a torná-los estruturados e de fácil acesso ao Cliente X. nas suas tomadas de decisão.

Após o desenvolvimento da camada ETL, nomeadamente estruturas hierárquicas construídas e dependências entre campos definidas, surgiu a necessidade de construir os relatórios dinâmicos de apoio à decisão, de acordo com os requisitos estipulados pelo Cliente X. (objetivos 6 e 7 – ver ponto 1.4.1 , página 8). Estes relatórios tinham como pressupostos: permitir uma interação total com os dados apresentados e com a forma como são apresentados, por intermédio de gráficos, tabelas, entre outros, permitindo a navegação dentro do painel e a modificação de filtros. Inicialmente, ficou definido um *template* que serviu como base para qualquer relatório construído pela equipa de desenvolvimento. Para cumprir todos os requisitos estipulados pelo cliente, foram realizadas várias reuniões periódicas onde foram detalhados os dados que iriam ser apresentados nos painéis interativos e qual a sua forma de apresentação. Depois de cada reunião, foram desenvolvidos os relatórios que constituíam a base dos painéis interativos e definidos os atributos com os nomes para os utilizadores finais e, também, as métricas necessárias para obter os resultados esperados pelo cliente, por intermédio da linguagem DAX.

Na definição e construção da camada de visualização foram tidos em consideração alguns aspetos e funcionalidades:

- Manter uma interface simples, com recurso a gráficos e tabelas sem grande variação do tipo e cores utilizadas;
- Permitir a definição de parâmetros (filtros) nas análises a realizar;
- Visualização de todos os indicadores operacionais relevantes num único relatório;
- Possibilidade de criar *Dashboards* simples e intuitivos que agregassem dados de múltiplas fontes e monitorizassem métricas de sucesso alinhados às estratégias de negócio;
- Informação disponível em plataformas e *apps online*;
- Gráficos evolutivos e comparativos;
- Ferramenta capaz de suportar a análise de grande volume de dados e com capacidade de atualização automática e envio/disponibilização imediata de informação.

A tecnologia de BI adotada pela Unipartner IT Services e Cliente X. e, que foi usada durante o estágio pela autora, foi o Microsoft *Power BI*, a qual permitiu a transformação de uma grande quantidade de informação fornecendo uma plataforma corporativa de alto desempenho, escalável e capaz de proporcionar visões profundas com relatórios ou *dashboards* interativos e funcionalidades analíticas para uso por uma grande quantidade de utilizadores através da *web* e dispositivos móveis.



### 3.6.1 Análise Preliminar de Requisitos

Concluída a fase anterior, foi necessário mostrar os resultados aos utilizadores finais. A exibição dos dados tratados e estruturados foi feito através do Microsoft *Power BI*, em detrimento das *pivots tables* usadas em Microsoft Excel, onde o utilizador obteve os resultados pretendidos através da aplicação das métricas dos grupos de medidas e dos valores das dimensões nos campos dos relatórios de Microsoft *Power BI*. Numa primeira fase, foi definida uma paleta de cores para implementar nos relatórios em *Power BI*.

Seguidamente, detalharam-se os requisitos solicitados pelo Cliente X. na construção dos relatórios:

- **Conformidade dos relatórios face aos requisitos definidos e de acordo com os dados provenientes dos relatórios mensais do Cliente X.:** os dados referentes aos campos das dimensões e das tabelas de Factos corresponderam aos dados presentes nas bases de dados operacionais. Os campos, a informação e os *layouts* de apresentação dos relatórios tinham de estar de acordo com o que foi estipulado nos requisitos e reuniões com o Cliente X.;
- **Layout e interface dos relatórios:** os relatórios do Cliente X. apresentaram no seu *layout*: logótipo da entidade bancária, com um título descritivo do relatório e filtros de pesquisa apropriados em todos os relatórios, nomeadamente, intervalos de datas definido que retornavam apenas registos cujo campo filtrado pertencesse ao intervalo e outros que o Cliente X. considerasse adequados para a sua pesquisa. Todos os campos de identificação de um relatório deveriam ter as mesmas medidas, tipos de letra e largura das colunas para se compreender a informação.

Esta fase iniciou-se com a criação de três relatórios pilotos em *Power BI* com relatórios iterativos com dados analíticos representativos dos elementos gráficos que atualmente apoiam a atividade e que se incluem na informação de gestão do Cliente X. no âmbito da atividade planeada (ciclo de vida e orçamento) e quotas, atividade corrente (suporte a clientes), nomeadamente no registo de pedidos de serviço e incidentes e na distribuição da atividade da empresa em horas por tipologia de atividade (planeada e corrente).

Como dados iniciais de análise, foram usadas fontes estruturadas de informação, sob a forma de tabelas de dados Excel, extraídas através de relatórios existentes, *queries* ou alimentados diretamente, sendo que não existiam gráficos ou quadros que tivessem sido tirados diretamente da solução. Depois, foi criada uma ligação dos modelos analíticos ao *Power BI*, no qual se prepararam *dashboards* ou relatórios em áreas de trabalho individuais ou em grupo. A publicação dos *dashboards* ou relatórios criados efetuou-se igualmente no portal de *Power BI*. Os relatórios criados constituíram a ferramenta de trabalho diária para os utilizadores atendendo ao âmbito de consulta de cada um.

O modelo de *Governance* para os relatórios de apoio à decisão assentou nos seguintes pressupostos:

- Os relatórios foram publicados na aplicação *Power BI* para consulta. O acesso às publicações não teve em conta níveis de permissões aplicados;
- As áreas de trabalho consistiram numa área individual (*My Workspace*) para utilizadores habilitados para a criação de visualizações, uma área de grupo (*Group Workspace*), por

grupo de trabalho, que foi utilizada para publicação das visualizações individuais em desenvolvimento e, uma área de grupo (*Group Workspace*) onde foram publicadas as visualizações para aprovação por parte de um departamento específico do Cliente X.

- A criação dos relatórios ocorreu na ferramenta *Power BI Desktop* enquanto a sua disponibilização ocorreu no *Workspace* e a sua visualização gráfica foi possível através da *App*.

Tendo em conta o exposto, surgiu a necessidade de encontrar uma solução mais eficiente e eficaz para a realização de relatórios de apoio à decisão que cumprissem as necessidades dos utilizadores finais e de estudar quais os efeitos dessa solução no desempenho do pessoal envolvido na produção dos relatórios. A produção de relatórios e *dashboards* que permitiram a visualização gráfica dos dados atuais e das tendências do negócio, por forma a simplificar a tomada de decisões, será referenciada no próximo ponto.

### 3.6.2 Construção e Implementação dos Relatórios

Nesta fase, foram construídos os relatórios e os *dashboards* que constituem um dos elementos de BI mais importantes, uma vez que permitem a transformação de conjuntos de dados em elementos gráficos. Define-se por relatório de BI, um componente que permita apresentar um conjunto de dados, que se façam representar por diferentes tipos de elementos, nomeadamente análises, gráficos e tabelas.

Com o objetivo de dar resposta a todas as necessidades que a solução apresentava, foi utilizado o *stack* do *Power BI*, ou seja, o serviço de *Power BI* e o *Power BI Desktop*, com o *layout* apresentado na Figura 16:

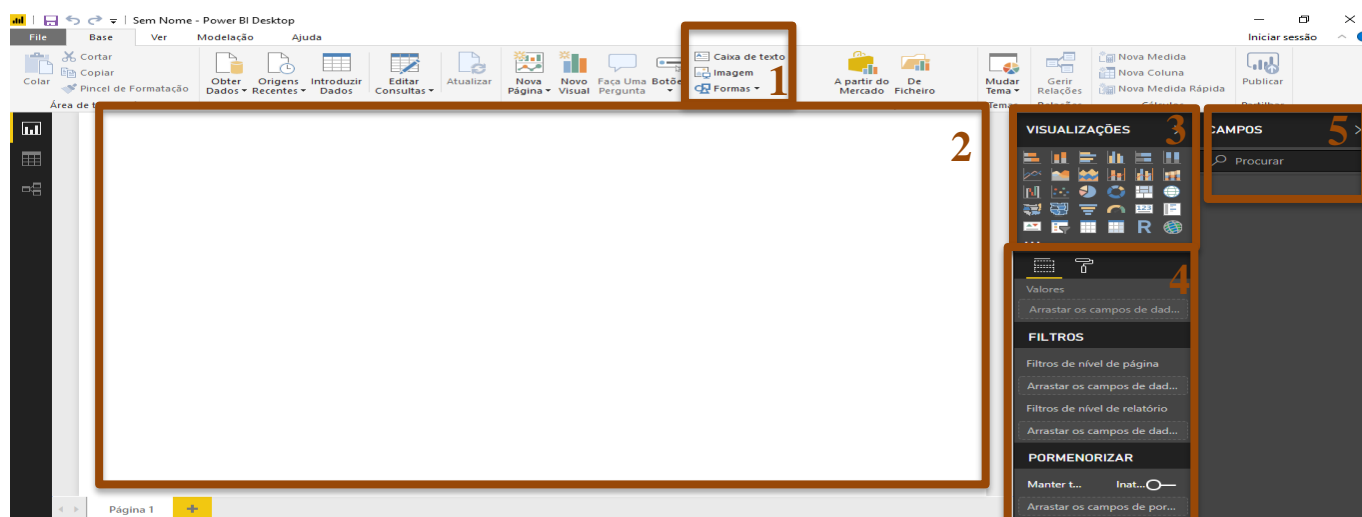


Figura 16: Área de Trabalho do Microsoft *Power BI*

Fonte: Elaborado pela autora

Legenda:

1. *Palette*: disponibiliza todos os elementos gráficos que podem ser adicionados a um relatório (por exemplo, imagens, campos de texto, formas);
2. Área de trabalho: zona que permite criar e pré-visualizar os relatórios;
3. Ferramenta de Estrutura (Visualizações): apresenta todos os elementos visuais e não visuais presentes nos relatórios;

4. Ferramenta de Propriedades: lista todas as propriedades da seleção ativa, permitindo a modificação das propriedades comuns a diversos elementos selecionados de uma só vez (por exemplo, tipo de letra, alinhamento do texto, cor de preenchimento, negrito, itálico, sublinhado) e definição de filtros de pesquisa;
5. Campos: onde se encontra definido o *Dataset*.

Cada um destes componentes desempenhou um papel distinto na solução, sendo o serviço de *Power BI* utilizado como área de trabalho corporativa e de disponibilização da informação e o *Power BI Desktop*, a ferramenta utilizada para o desenvolvimento de novos relatórios.

A arquitetura da solução, bem como as metodologias propostas foram criadas assentes no pressuposto de que existiam três grupos de utilizadores que deveriam ser criados ao nível da AD, para organização interna e não para definições de perfis ao nível do *Power BI* e dizem respeito ao conjunto de utilizadores que constroem novos relatórios e aos que apenas consomem os relatórios já criados:

- *Reports-Viewers*: grupo de utilizadores que consomem informação da solução e cujo acesso ao *Power BI* é feito via *App*;
- *Reports-Publishers*: grupo de utilizadores que criam e publicam relatórios, cujo acesso ao *Power BI* é feito via *Workspace* e *App*;
- *Self-Service*: grupo de utilizadores que criam relatórios, cujo acesso ao *Power BI* é via *Workspace Self-Service*.

Inicialmente, definiu-se o *layout* do relatório de acordo com os requisitos que são transversais a todos os relatórios, dos quais se destacam (Figura 17): inclusão do logótipo do cliente (1): por questões de confidencialidade, foi omitido o logótipo do Cliente X.; data de atualização do relatório (2); identificação dos relatórios: Título do relatório e fonte correspondente (3); possibilidade de filtrar os dados através de um conjunto de filtros definidos para cada relatório (4); os campos apresentados deveriam ter um cabeçalho que os identificasse (5).



Figura 17: *Layout* definido para os relatórios em *Power BI*  
Fonte: Elaborado pela autora

Na entrega da solução, foi disponibilizado um relatório sobre os dados da fonte de dados A, B e C, bem como os seus respetivos *datasets*, que constituem o ponto de partida para a construção de um novo relatório. Posto isto, a construção dos novos relatórios foi feita sobre este *dataset*, isto porque este tinha configurado o modelo de dados e todas as métricas associadas. Desta forma, a construção de novos relatórios foi realizada através do *Power BI Desktop*, utilizando a ligação aos serviços online – *Power BI service* – e, selecionando o *workspace* e o *dataset* pretendidos. Este novo relatório ficou com uma ligação direta para o *dataset* escolhido, permitindo a gestão dos relatórios centralizada, ou seja, ao ser criada uma nova métrica esta era automaticamente refletida para todos os relatórios construídos sobre este *dataset*. Após a ligação do novo relatório ao *dataset* e a construção de todas as análises pretendidas, o mesmo foi publicado para o *Workspace*, ficando disponível para toda a equipa de desenvolvimento e também para consumo dos restantes utilizadores após atualização da *App*. A gestão do conteúdo

existente neste *Workspace* e disponibilizado na *App* (consumo de relatórios) foi garantido pela equipa responsável do Cliente X. O consumo de relatórios foi feito com recurso à *App*. Cada *App* só podia ser acedida por elementos que pertencessem a um grupo AD definido, por intermédio de uma conta no Office 365, única para cada elemento. Após a criação da *App*, foi necessário adicionar o respetivo grupo de utilizadores de consumo de relatórios à aplicação e, a partir desse momento, todos os utilizadores que estivessem inseridos nesse grupo passavam a ter acesso à aplicação e a todo o seu conteúdo.

Um dos objetivos delineados anteriormente era permitir que qualquer utilizador pudesse desenvolver novos relatórios com as análises que pretendesse e, assim, enriquecer o *reporting* da organização. De forma a alcançar este objetivo, a solução incluiu um *workspace* de desenvolvimento, totalmente livre, onde qualquer utilizador era livre para construir, editar e eliminar qualquer relatório (*self-service BI*).

A partir do desenvolvimento da camada de ETL, começaram-se a obter dados para a realização dos relatórios operacionais no Microsoft *Power BI*, por ser fácil de usar, funcional e ter bom desempenho. Por não depender da existência de dados e por se conhecerem os campos necessários, foi possível avançar com o desenho dos relatórios, durante o desenvolvimento do processo de ETL. A equipa de desenvolvimento, neste ponto, foi reforçada com a chegada de um terceiro elemento, conhecedor da dinâmica de construção de relatórios em Microsoft *Power BI*. Ainda na parte da implementação do *dataset* com os dados de análise no *Power BI*, foi possível a definição das hierarquias mediante as escolhidas na modelação dimensional. As hierarquias possibilitaram a utilização da opção de “*drill*” que permitiu aos utilizadores navegar nas hierarquias das dimensões, ou seja, do elemento mais elementar até ao mais agregador e vice-versa.

Nas figuras seguintes, encontra-se o resultado final dos relatórios operacionais correspondentes às duas fontes de dados operacionais (A e B). No caso da **fonte de dados A**, a produção e disponibilização de relatórios teve em conta duas abordagens: uma página para os “Pedidos em Carteira” (Figura 18) que refletissem análises mais antigas e uma segunda página para “Pedidos Abertos e Fechados” (Figura 19) que refletissem o número de pedidos abertos e fechados atualmente:



Figura 18: Página 1 do Relatório da Fonte de dados A – “Pedidos em Carteira”  
Fonte: Elaborado pela autora

Na Figura 18, foi possível analisar todos os “Pedidos em Carteira” que correspondessem ao somatório dos pedidos no estado “Em Análise” e “Em Execução”, tendo em conta o “Sub-sub estado”, “Data de Criação” desses pedidos, “Origem” e “Dimensionamento”. Nesta página, foi, ainda, possível, filtrar a informação, recorrendo aos seguintes filtros: “Tipo de Pedido”, “Entidade”, “Gestor de Projecto”, “Estrutura”, “Data de Fotografia”. Em todos estes filtros, foi possível seleccionar todas as opções ou uma apenas uma.

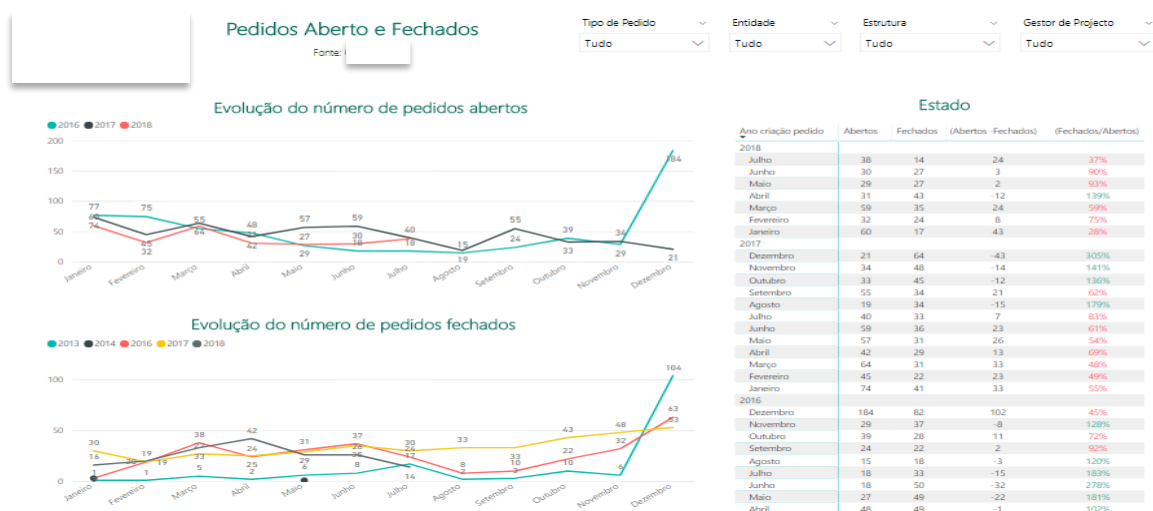


Figura 19: Página 2 do Relatório da Fonte de dados A – “Pedidos Abertos e Fechados”  
Fonte: Elaborado pela autora

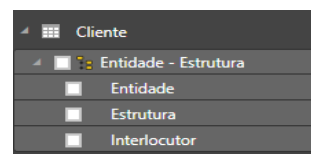
Na Figura 19, foi possível analisar todos os “Pedidos Abertos e Fechados”, ou seja, a evolução do número de “Pedidos Abertos”, tendo em conta o “Ano de Criação” e, com possibilidade de *drill-down*, para mês e dia e, a evolução do número de “Pedidos Fechados”, tendo em conta o ano, mês e dia de fecho. Ainda, foi possível analisar estes dados, tendo em conta o “Estado” do pedido, nomeadamente o “Sub Sub-Estado”, por análise da “Data de Criação”, “Entidade” e “Estrutura”. Os filtros de página escolhidos foram os seguintes: “Tipo de Pedido”, “Entidade”, “Estrutura” e “Gestor de Projecto”. Em todos estes filtros, foi possível seleccionar todas as opções ou apenas uma.

Ainda em relação a este relatório, foram definidas métricas: diferença entre pedidos abertos e fechados e relação entre pedidos fechados e abertos, através da linguagem DAX, como se pode ver de seguida:

(Abertos - Fechados) =  
[Nº de Pedidos] - [Fechados]

(Fechados/Abertos) =  
DIVIDE([Fechados]; [Nº de Pedidos])

Quanto às hierarquias, estas apareciam representadas pela imagem do lado direito, para o exemplo do “Cliente” (Entidade > Estrutura > Interlocutor):



No caso da **fonte de dados B**, a produção e disponibilização de relatórios teve em conta duas abordagens: uma página para os “Projectos” (Figura 20) que refletissem as análises para os projetos desenvolvidos pelo Cliente X. e uma segunda página para “Horas Reportadas” (Figura 21) que refletissem as horas reportadas nos projetos, de acordo com várias análises. Em concordância com o cliente, a equipa de desenvolvimento, para a fonte de dados B apresentou o relatório referente aos “Projectos” numa visão de atividade planeada.

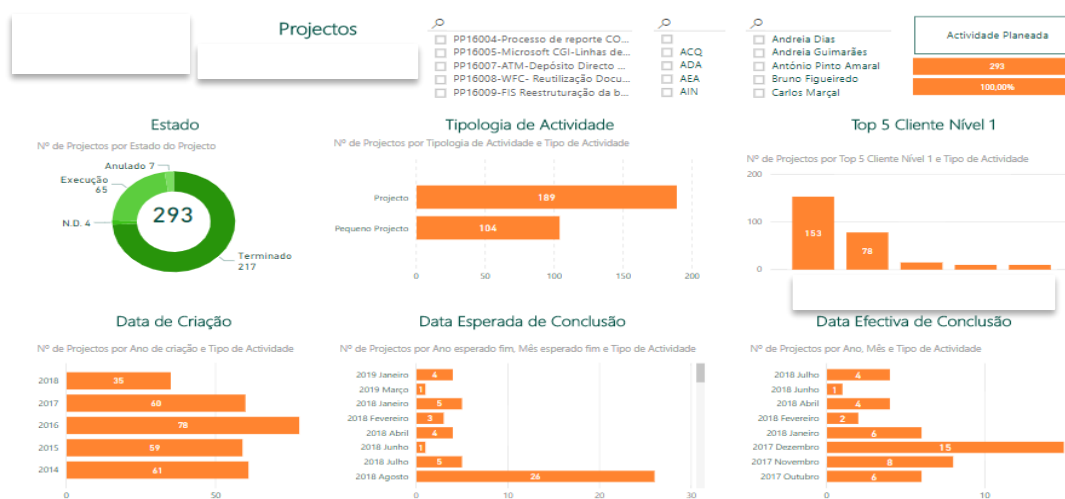


Figura 20: Página 1 do Relatório da Fonte de dados B – “Projectos”

Fonte: Elaborado pela autora

Na Figura 20, foi possível analisar todos os “Projectos” que fazem parte do âmbito da fonte de dados B, tendo em conta o “Estado”, “Tipologia de Actividade”, “Top 5 Cliente Nível 1”, “Data de Criação”, “Data Esperada de Conclusão” e “Data Efectiva de Conclusão”. Nesta página, os filtros

escolhidos foram os seguintes: “Nome do Projecto”, “Cliente Nível 2” e “Gestor de Projecto”. Em todos estes filtros, foi possível seleccionar todas as opções ou uma determinada opção e pesquisar manualmente a informação pretendida, através da opção da lupa.

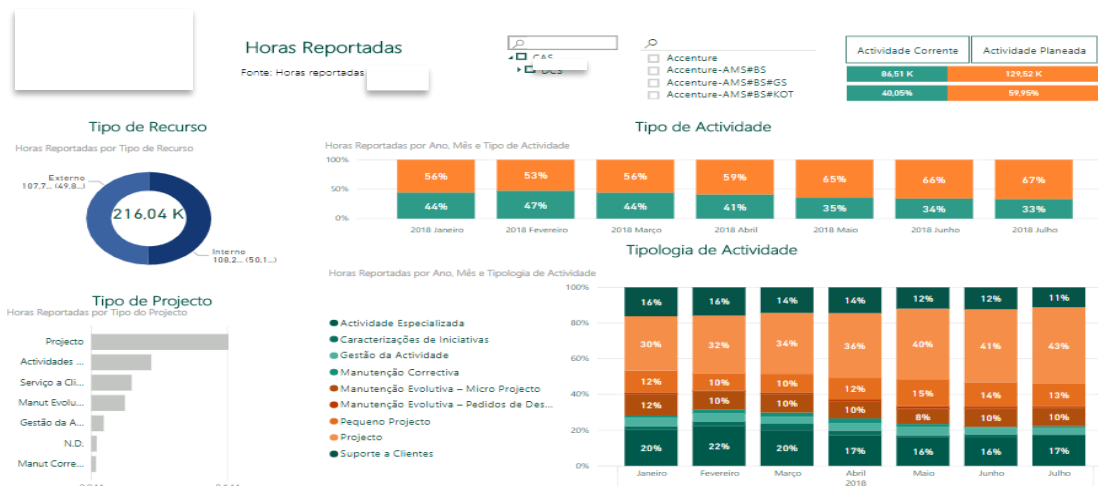


Figura 21: Página 2 do Relatório da Fonte de dados B – “Horas Reportadas”

Fonte: Elaborado pela autora

Na Figura 21, foi possível analisar “Horas Reportadas” no âmbito da fonte de dados B, tendo em conta o “Tipo de Recurso”, “Tipo de Actividade”, “Tipo de Projecto” e “Tipologia de Actividade”. Nesta página, foi ainda possível filtrar informação, recorrendo aos filtros: “Hierarquia RBS”, “Empresa do Recurso” e “Tipo de Actividade”, diferenciada em “Actividade Corrente” e “Actividade Planeada”. Em todos estes filtros, foi possível seleccionar todas as opções ou uma determinada opção e pesquisar manualmente a informação pretendida, através da opção 🔍 para os dois primeiros filtros.

Foi definido, ainda, um *tooltip* com a designação “Distribuição do número de projectos por Estado no tipo de projecto” na tabela de Factos “Projectos” recorrendo à linguagem DAX:

```
ToolTipName = CONCATENATE("Distribuição do número de projectos por Estado no tipo de projecto: "; Projecto[Tipologia de Actividade])
```

Esta funcionalidade permitiu ter acesso a uma análise complementar para uma determinada análise identificada.

Os *dashboards* de BI permitiram visualizar os dados de uma forma diferente da que foi disponibilizada pelos relatórios de apoio à decisão. Todavia não foram mostrados neste projeto, por serem propriedade de um dos responsáveis do departamento que requisitou o projeto. Este colaborador do Cliente X. ficou encarregue de criar os *dashboards* a partir dos relatórios realizados, tendo em conta as dimensões e as métricas definidas no *dataset* de cada fonte de dados e, fazendo uso do conceito de *self-service BI*.

O Portal de *Reporting* Corporativo foi desenvolvido com acesso a todas as funcionalidades residentes no universo do Microsoft Power BI (*Workspace*, *App* e *Power BI Desktop*). Foi definida, com a colaboração do Cliente X., a estrutura do Portal nomeadamente em termos de acessos e da nomenclatura das áreas onde foram disponibilizados os relatórios e *dashboards Power BI Desktop*.

Nome do Projecto	Horas Reportadas
PR17041-IFRS9 - Implementação	1.908.00
PR18015-Evolução dos Sistemas Aplica...	1.624.50
PR18019-Reformulação Oferta Crédito ...	1.345.25
PR18002-i9 - Desenho do Modelo e Re...	1.213.75
PR17034-Reformulação da CRC - Fases...	881.50
PR15044-Focus ALM	686.00

### 3.6.3 Validação do trabalho realizado

A validação dos relatórios operacionais foi concretizada por elementos da equipa funcional do Cliente X. bem como por elementos da equipa técnica de desenvolvimento da Unipartner IT Services. Os testes efetuados abrangeram a interface utilizada (Microsoft Power BI), o desempenho e os dados presentes nos relatórios, por confrontação com a prova de conceito que continha os relatórios mensais, inicialmente disponibilizado pelo cliente.

Essencialmente, os erros detetados foram: falta e incorreção de alguns campos de dados, a não apresentação de determinados filtros de pesquisa na área de trabalho dos relatórios e melhorias na definição de gráficos a utilizar. Uma vez que foram definidas múltiplas métricas, verificou-se que estas foram corretamente apresentadas nos relatórios e testou-se o funcionamento dos filtros que as associavam a outros elementos, por forma a garantir a coerência de informação nos diferentes gráficos.

Na prática, a validação destes relatórios ocorreu nas reuniões formais já previstas, ou de forma mais livre, quando necessário, garantindo que os resultados finais forneciam a informação pretendida.

## 3.7 Plano Efetivo: Análise aos desvios do Planeamento do Projeto

Nesta parte do projeto, é necessário mencionar, em linhas gerais, a decomposição, atribuição e calendarização das tarefas realizadas e uma explicação da diferença entre o planeado *vs* o executado, analisando as razões dos desvios ocorridos. De uma forma geral, o plano geral deste estágio dividiu-se em três grandes blocos: *self-study*, projeto 1 com uma entidade pública e o projeto 2 com uma entidade bancária, o Cliente X. O não seguimento do Projeto 1, partiu do orientador da Unipartner IT Services, por considerar ser uma mais valia para o percurso da autora em BI, poder acompanhar um projeto desde o início e, por considerar ser mais enriquecedor para a construção do trabalho de projeto da autora.

O diagrama de Gantt apresentado em apêndice, encontra-se organizado por tarefas principais e respetivas sub-tarefas, sendo a sua duração medida em intervalos de tempo, em dias. O plano de trabalho efetivo apresenta algumas diferenças em relação ao do relatório preliminar devido à mudança do projeto, sendo redefinidos alguns capítulos.

Para que haja um conhecimento detalhado e transparente do trabalho realizado ao longo dos nove meses de estágio, é indispensável a apresentação das justificações das várias alterações dos planos elaborados numa fase anterior, e a apresentação da calendarização de todo o trabalho realmente efetivado. A partir da mudança, de um modo geral, o projeto decorreu conforme planeado, havendo, no entanto, algumas exceções. Ao contrário do que estava planeado, as tarefas previstas para a primeira fase de entrega ao Cliente X., foram incluídas numa segunda fase, contribuindo para a extensão da duração do projeto.

Numa primeira fase (Plano inicial definido):

- Levantamento dos requisitos funcionais;
- Estudo das tecnologias, metodologias e trabalhos relacionados existentes;
- Definição dos modelos de dados;
- Validação dos dados presentes nos relatórios mensais da prova de conceito;
- Correção de possíveis erros encontrados nos relatórios mensais da prova de conceito;
- Segurança: estudo dos mecanismos de segurança, implementação e testes;
- Escrita do relatório preliminar.



Na segunda fase (Plano efetivado):

- Definição dos modelos de dados;
- Validação dos dados presentes nos relatórios mensais da prova de conceito;
- Correção de possíveis erros encontrados nos relatórios mensais da prova de conceito;
- Segurança: estudo dos mecanismos de segurança, implementação e testes;
- Implementação e desenvolvimento do processo de ETL;
- Construção do DW;
- Produção dos relatórios em *Power BI*;
- Validação dos relatórios;
- Finalização do Relatório Final do Projeto;
- Preparação da apresentação do Projeto.

O relatório preliminar foi entregue apenas no dia 5 de janeiro de 2018, referenciando somente o primeiro projeto em que a autora esteve envolvida. Esta entrega aconteceu, sensivelmente, dois meses depois do início efetivo do projeto (6 de novembro de 2017). No primeiro projeto, a autora apenas acompanhou a realização de trabalhos, uma vez que a sua contribuição mais notória acontecia apenas na fase 3 e 4 desse projeto, com início a 9 janeiro de 2018. Posto isto, foi constatado que não havia a necessária sustentabilidade física que justificasse a autora continuar a fazer parte do projeto 1 e, por isso, o presente documento aborda apenas todo o trabalho desenvolvido no projeto 2, albergando alguns conceitos comuns ao projeto 1. A **mudança de projeto** aconteceu no dia 1 de janeiro de 2018, sendo que houve também mudanças dos planos gerais e detalhados do mesmo. Neste novo projeto, foi criado um *reporting* corporativo para o Cliente X. de modo a automatizar a criação de relatórios de apoio à decisão. Esta mudança foi proposta pelo orientador da empresa e comunicada ao orientador da faculdade e constituiu uma mais valia para a autora, uma vez que lhe permitiu assistir e participar na construção de um sistema de base de dados e todos os conceitos inerentes ao universo de BI. Tendo um projeto alternativo, foi alterado o âmbito do projeto, o plano geral e detalhado, os pressupostos sobre as quais assenta o projeto, bem como a contribuição da autora no projeto que, a nível do projeto 2, foi a 100%.

Na fase de validação de dados, houve um atraso de três semanas por causa de dificuldades de agenda do Cliente X. Também houve alteração das hierarquias de dados dentro da empresa, o que levou a ajustes na estrutura de algumas das dimensões criadas, nomeadamente, “Tipo de Cliente” para a fonte de dados A, o que exigiu várias reuniões de redefinição, com impacto na calendarização do projeto, em quase quatro semanas. As atividades de definição da arquitetura da solução e questões de segurança de dados bem como a realização de testes sofreram um desvio de algumas semanas, sensivelmente cinco, em relação ao plano inicial, por estarem dependentes de validação por parte do Cliente X., atrasando a continuação dos trabalhos.

Finalizando, a duração de 21 semanas, inicialmente definida para a concretização do projeto, foi adiada para um total de 33 semanas, aproximadamente 8 meses. Face a todos estes atrasos, a autora concluiu todas as tarefas não dependentes entre si, sendo destacada, algumas semanas, para auxiliar noutros projetos até à retoma de trabalhos do projeto em estudo.

A Tabela 14 é referente ao período proporcionado pela Unipartner IT Services para elaboração dos respetivos relatórios, preliminar e final. Enquanto a coluna da direita se refere ao tempo disponibilizado pela empresa para a redação dos relatórios, a coluna à esquerda mostra o tempo despendido na redação dos documentos por parte da autora. Pode-se constatar que foi necessário mais tempo para a realização destes documentos, pois em paralelo, a autora estava a 100% na concretização

do projeto e, ao mesmo tempo, estava envolvida noutros projetos que exigiam esforço e tempo fora do horário de trabalho previsto.

Tabela 14: Calendarização da realização dos relatórios preliminar e final

Relatórios	Tempo disponibilizado pela Unipartner IT Services para a concretização dos relatórios	Duração Efetiva da Concretização da Tarefa
Relatório Preliminar	30 dias	15 dias
Relatório Final	90 dias	90 dias

Fonte: Elaborado pela autora

## 3.8 Sumário

Neste capítulo, foram apresentados os seguintes tópicos: os requisitos e objetivos do problema exposto pelo Cliente X.; o ambiente de trabalho onde decorreu o estágio; a organização do trabalho em equipa, o processo de desenvolvimento de *software* adotado pela equipa; a caracterização dos dados do processo de negócio, segundo as três fontes de dados operacionais; a metodologia de trabalho aplicada; a descrição, em linhas gerais, das tecnologias, ferramentas e linguagens utilizadas; a definição da arquitetura de solução; a explicação do desenvolvimento dos processos de ETL e construção do DW; e, por fim, o tratamento da camada de visualização (*dashboards* e relatórios em *Power BI*). Também neste capítulo, analisou-se os desvios que ocorreram em relação ao plano inicialmente definido.



# Capítulo 4

## Conclusão

### 4.1 Considerações Finais

Cada vez mais os agentes de decisão necessitam que lhes sejam disponibilizadas ferramentas analíticas que os auxiliem no seu processo de tomada de decisão. Tradicionalmente, os Sistemas de DW e os Sistemas de Processamento Analítico apresentam inúmeras vantagens quando se pretende analisar informação e processar grandes volumes de dados. Os sistemas de informação contribuem para o sucesso das empresas, podendo disponibilizar informações rapidamente, ajudando os gestores a analisar os seus dados de forma a perceberem o que está errado no negócio. Desta forma, poderão aumentar a satisfação dos clientes e, conseqüentemente, alcançar o aumento dos lucros. As empresas têm que optar por uma postura de trabalho mais voltada à gestão da informação para poderem criar estratégias que atendam melhor aos seus clientes e coloquem a empresa num patamar de competitividade mais lucrativo. Como o mercado está cada vez mais competitivo, as empresas e clientes passam a ser mais ágeis e exigentes. Torna-se muito difícil para a empresa conseguir sobreviver sem qualquer ferramenta de BI, fazendo com que a sua implementação lide com uma visão objetiva e correta dos negócios de forma a facilitar a tomada de decisão, visando vantagens competitivas. São utilizados métodos, técnicas, ferramentas e tecnologias para auxiliar a conceção, criação, gestão e análise de processos de negócio para que estes se tornem mais eficientes (reduzindo os recursos utilizados), eficazes (produzindo os resultados esperados) e adaptáveis (atendendo às necessidades do mercado).

A presente iniciativa decorre de um trabalho autónomo de natureza profissionalizante e consistiu na alteração do modelo de construção de relatórios, que resultou da análise de três fontes. Partindo deste fator, foi elaborado um diagnóstico prévio que identificou a necessidade de definir e implementar uma solução de BI de modo a controlar melhor os serviços prestados pelo Cliente X. O rápido desenvolvimento e evolução da informática nos últimos anos, trouxe o aparecimento de BI e a sua contribuição para um melhor alcance de resultados nas empresas.

Este trabalho propôs o desenvolvimento e a implementação de um DW corporativo, contribuindo para que a empresa fornecesse ao seu cliente, produtos e serviços mais personalizados e em conformidade com as necessidades do mesmo, através da organização dos dados. Como produto final, foram disponibilizados, para as três fontes de dados, relatórios em *Power BI* dinâmicos com um modelo de dados organizado e um *layout* intuitivo.

Neste ponto, são descritas as principais conclusões retiradas aquando da realização do projeto na Unipartner IT Services, destacando as principais contribuições, competências adquiridas e apreciação crítica, principais dificuldades e limitações e sugestões para melhorias na solução implementada e trabalho futuro.

## 4.2 Principais Contribuições

A integração da autora na área de BI, na equipa de trabalho e no ambiente da empresa ocorreu de forma natural e foi muito positiva. Relativamente à realização do trabalho proposto para este projeto, a autora demonstrou iniciativa e capacidade na proposta de alternativas e soluções.

Neste final de estágio, a autora ficou a compreender o processo de implementação de uma solução de BI, no seu todo. No decorrer deste estudo, a autora constatou que a produção de relatórios é, muitas vezes, realizada de uma forma não automática e que os colaboradores da empresa estão, na sua grande maioria, recetivos à automatização do processo. Contudo, não querem perder o controlo da informação apresentada, sendo importante conseguirem realizar alterações à informação ou estrutura dos documentos, os quais devem ser o mais personalizáveis possível.

Houve um acompanhamento constante por parte dos elementos da equipa em que a autora esteve integrada no decorrer do estágio, mostrando-se sempre disponíveis a responder a qualquer dúvida que surgisse. Isto incluiu atividades como formação, alinhamento dos objetivos das unidades de negócio, assim como a promoção de uma conduta adequada.

A autora, tal como fora referido, integrou a equipa de projeto a 100% acompanhando o projeto desde o processo de levantamento de requisitos de forma concisa e correta para que o processo de tomada de decisão fosse o mais correto e coerente, passando pela definição das métricas de análise do negócio e terminando na produção de relatórios, principalmente da fonte de dados A. Ao longo da realização do projeto outras tarefas ficaram ao encargo da autora: análise de dados incorretos, incompletos, desatualizados ou inacessíveis da fonte de dados A em conjunto com os responsáveis por esta fonte.

De uma forma geral, após uma reunião inicial com o representante do cliente para apurar o âmbito do projeto, seus problemas e objetivos, os passos foram os seguintes:

- Análise de Requisitos – reuniões constantes para averiguar e perceber o que realmente faria sentido no âmbito do projeto, decisões a serem tomadas e principais funcionalidades do *reporting* corporativo. Consulta do estado atual do sistema de elaboração de relatórios, fazendo a avaliação do que têm para o que devem e podem ter, simplificando e não alterando demasiado a composição dos dados (objetivo 1 – ver ponto 1.4.1 , página 8);
- Implementação dos processos de ETL no SSIS e construção do DW (objetivos 2, 3, 4 e 5 – ver ponto 1.4.1 , página 8);
- Desenho e produção dos relatórios da fonte de dados A e B, e acompanhamento da construção dos relatórios da fonte de dados C. Os relatórios construídos são funcionais e cumprem os requisitos definidos pelo Cliente X. (objetivos 5 e 6);
- Registo documental da estrutura do *reporting* corporativo, das soluções encontradas e implementação das mesmas. A autora, neste ponto, ficou encarregue da elaboração dos documentos funcionais das três fontes de dados;
- Apresentação e discussão da solução provisória ao cliente – *workshops* funcionais e reuniões da apresentação de propostas e da solução final ao Cliente X. (objetivo 6 – ver ponto 1.4.1 , página 8).

Neste projeto, autora tentou sempre dar o seu melhor e fazer um bom trabalho, apresentando sugestões e opiniões sempre que necessário, colocando perguntas pertinentes para atingir os objetivos profissionais.

Este estágio possibilitou a aquisição de competências técnicas específicas, por intermédio da inclusão da autora numa equipa multidisciplinar *Scrum* e por ser exigido à autora documentar, de acordo com as normas da empresa, todos os passos definidos para a concretização da solução.

A redação dos dois relatórios foi deixada para último lugar por falta de tempo, mas como a qualidade é o mais importante, a preferência recaiu sobre prolongar a escrita do trabalho de projeto para garantir um melhor resultado.

## 4.3 Competências Adquiridas e Apreciação Crítica

Com este trabalho de projeto, foi possível integrar conhecimentos e adquirir competências ao longo do curso do projeto tendo em vista a apresentação de soluções ou recomendações sobre o problema de *reporting* de dados de que o Cliente X. era alvo. Foram valorizadas as dimensões de carácter multidisciplinar e experimental, sem esquecer a necessidade de um enquadramento teórico e justificação metodológica. A metodologia seguida pelo projeto foi feita com base em literatura com conceitos e implementações padrão para a construção de um *reporting* corporativo e numa abordagem técnico-científica.

A realização de um estágio na área de BI torna-se benéfica a todos os níveis, quer pela experiência profissional adquirida, quer pela possibilidade de interagir dentro de uma organização com uma elevada componente tecnológica. Esta oportunidade única permitiu à autora adquirir e consolidar conhecimentos práticos e teóricos na área de BI. As atividades desenvolvidas no decorrer do estágio consistiram na construção de uma solução de BI, para colmatar as falhas existentes na produção de relatórios por parte de uma entidade bancária e, na elaboração de relatórios finais para vários departamentos do Cliente X. O estágio permitiu adquirir experiência numa tecnologia de referência no mercado de BI (Microsoft *Power BI*) e aprofundar os conhecimentos acerca da atividade de uma entidade bancária. Por vezes, a mudança de costumes é uma barreira para os utilizadores, tornando-se difícil abandonar hábitos que foram adquiridos ao longo do tempo, em que se familiarizam com determinada aplicação. Neste projeto, esta situação foi facilmente contornada, pela avaliação da adesão e anseio pela mudança e pela novidade.

A realização deste projeto trouxe vários benefícios a nível pessoal: integração numa instituição externa; capacidade de aprendizagem; capacidade técnica; autonomia; capacidade de trabalho em equipa; iniciativa; visão abrangente; capacidade de comunicação (verbal e escrita); e, capacidade de organização e novas competências em soluções tecnológicas (*SQL Server*, *SSIS* e *Microsoft Power BI*). Foram cumpridas metas pessoais e profissionais, destacando-se as seguintes: aplicabilidade da solução para a realidade do cliente; cumprimento dos objetivos definidos para o projeto; desafios de implementação, desenho tecnológico e atividades de implementação; documentação referente ao projeto; instalação e configuração de serviços e aplicações; análise de problemas apresentando abordagens/soluções inovadoras para os desafios identificados; experiência e *Know-how*: conhecimentos das regras de negócio, melhores práticas, implementações de soluções de BI e conhecimento especializado de arquiteturas de sistemas de informação para sustentar a extração e normalização de informação a partir dos sistemas operacionais.

De uma maneira geral, este estágio constituiu uma excelente forma de exposição à realidade empresarial, através da integração da autora como elemento produtivo, em projetos, com oportunidade de adquirir experiência nas áreas de conhecimento e ferramentas adotadas e tomar consciência das dinâmicas próprias da atividade de consultoria e das metodologias de acompanhamento, controlo e gestão do projeto.

Através do nível de desempenho no decorrer do projeto, houve um despertar notório quanto às competências e técnicas da autora. Foi extremamente gratificante, para a autora, a participação neste projeto inserido na área de desenvolvimento e poder integrar uma equipa de projeto júnior e sénior, com muita experiência nas TIs, na área de atuação no setor bancário. Também foi importante desenvolver o projeto nas instalações do cliente de forma a ter um contato mais direto e mais participativo.

O facto de ter havido reuniões diárias para apresentação do trabalho feito até então, esclarecimento de dúvidas face a quaisquer questões relacionadas e *feedback* respetivo, contribuiu para o bom sucesso do projeto. Contudo, a comunicação ao longo deste projeto foi uma componente muito importante pois os documentos e informação passada, não eram os mais atuais, sendo que houve a necessidade de validação constante com o cliente que, por diversas razões, não foi uma tarefa assim tão linear. Este não tinha muita disponibilidade, pois tinha em mãos outros projetos, para além do seu trabalho do dia a dia.

A autora foi capaz de resolver todos os desafios propostos, resolvendo-os da melhor maneira possível com o tempo disponível. Durante todo o trabalho desenvolvido, houve necessidade de um esforço suplementar de aprendizagem visto que a autora não estava familiarizada na área em que o estágio foi inserido, tendo sido imediatamente um desafio em termos de aprendizagem e descoberta de novas tecnologias, ferramentas e linguagens. A tarefa mais complicada foi o desenvolvimento dos processos de ETL, para além de ter sido a mais demorada.

Fazendo uma autoavaliação pode-se afirmar que as tarefas realizadas durante o estágio foram positivas pois os projetos foram concluídos com sucesso e dentro dos prazos previstos. Este estágio permitiu a integração num ambiente empresarial, o qual era totalmente desconhecido, mas que se revelou bastante interessante tendo sido uma ótima experiência, tanto a nível pessoal como profissional. Este estágio além de ter utilizado conhecimentos adquiridos ao longo do mestrado, foi uma constante aprendizagem pois a autora trabalhou numa área da informática onde nunca tinha estado envolvida.

Em conclusão, os produtos resultantes endereçaram de forma direta os objetivos propostos, aumentando a eficiência, eficácia e efetividade dos serviços prestados internamente pelo Cliente X. na realização de atividades comuns de produção de relatórios de forma contextualizada nas atividades das entidades protocoladas do Cliente X. De modo a assegurar que a manutenção futura das soluções será feita de forma autónoma, houve, ainda, transferência de conhecimento para os responsáveis das equipas internas.

## 4.4 Principais Dificuldades e Limitações

O desenvolvimento do projeto teve algumas situações que dificultaram a normal progressão dos trabalhos a um ritmo constante. Primeiro, as principais dificuldades sentidas na realização do projeto surgem da utilização de algumas tecnologias, nomeadamente, *SQL Server* e *Microsoft Power BI*, uma vez que o contexto académico é totalmente diferente do contexto empresarial e porque a autora ainda se encontrava num nível elementar de utilização das mesmas. Segundo, foi difícil, no início, a interiorização da perspetiva do negócio tendo em conta as necessidades do cliente. Terceiro, por se tratar de um projeto construído de raiz, foi complicada a definição de um plano de trabalho, delimitar o início,

etapas seguintes e conclusão do projeto. Quarto, a leitura e a reestruturação das fontes de dados (seleção e distinção entre informação relevante e não relevante) tornou o processo moroso no início do projeto. Apesar da autora desenvolver a sua atividade nas instalações do cliente, a sua disponibilidade nem sempre era constante. Quinto e último, prendeu-se com a definição de datas acordadas e atribuição de trabalhos a cada um dos intervenientes da equipa técnica da Unipartner IT Services.

Como qualquer projeto no setor público ou privado, é sempre um desafio conseguir obter toda a informação necessária e no tempo certo para apresentar uma solução rápida e eficaz que ultrapassasse todos os obstáculos.

Em todas as fases de construção do DW corporativo, surgiram dificuldades, nomeadamente:

- Conhecimento do negócio e análise das fontes de informação disponíveis: a autora desconhecia totalmente o setor bancário, tendo este projeto sido uma mais-valia no conhecimento de uma nova realidade;
- Definição de um plano de trabalho (onde começar e pelo quê? Que etapas se seguem? E falta de disponibilidade por parte do cliente);
- Definição e implementação de um modelo multidimensional: a principal dificuldade na perceção das relações existentes entre os campos pertencentes às diferentes fontes de dados e as tabelas de facto que tinham de ser criadas;
- Definição de um processo de ETL automatizado: na construção do processo de ETL surgiram várias dúvidas na sintaxe e construção de *queries* da linguagem SQL e na criação de um *job* final que integrasse tudo o que fora implementado;
- Na construção dos relatórios em Microsoft *Power BI*, a principal dificuldade surgiu na manipulação da linguagem DAX, na definição de métricas.

Apesar do rigor nos procedimentos empregados, o presente projeto apresentou algumas limitações. Salientam-se as mais importantes: falta de disponibilidade do cliente para reuniões de esclarecimento de dúvidas do projeto e validação de requisitos. Por ser um cliente da banca, os mecanismos de segurança a implementar em cada projeto estiveram sempre dependentes de regras de *firewall* que por si só, exigem tempo e recursos.

A experiência profissional permitiu retirar várias conclusões sobre o mundo profissional e o seu funcionamento obtendo uma balança de pontos positivos e menos positivos. Pelo lado mais positivo, temos a integração no mercado de trabalho por um período mais longo do que o já obtido noutra experiência, participação em projetos variados e a integração e colaboração com várias equipas. A contribuição de melhoria de negócio nas várias empresas, as competências adquiridas e a experiência em si também foram pontos gratificantes extraídos do trabalho. Como aspetos menos positivos apresenta-se a consolidação da ideia inicial de que as empresas não investem muito na formação dos seus colaboradores; a comunicação limitada com o cliente e a transparência da informação com a equipa de desenvolvimento sobre alterações no projeto ou decisões discutidas em reuniões; e a constante pressão para se realizar um trabalho bem executado e de uma forma rápida.

Em suma, na continuação do trabalho futuro, estas lacunas serão colmatadas por uma abordagem mais teórica das ferramentas, tecnologias e linguagens a utilizar, possibilitando o bom curso do trabalho.



## 4.5 Trabalho Futuro

A realização deste projeto teve como meta a otimização da criação dos relatórios por parte do cliente, devido à melhoria do acesso a dados. O objetivo inicialmente proposto foi concretizado – a produção de um DW corporativo – que criasse as condições necessárias para a criação de relatórios em *Power BI* em substituição da produção manual de relatórios no Microsoft Excel.

No entanto, muitas melhorias ainda ficaram por implementar. Considerando os benefícios proporcionados neste trabalho, os seguintes tópicos podem ser candidatos de um estudo em maior profundidade e apresentados em trabalhos futuros, numa perspetiva de manutenção evolutiva:

- Implementação da solução criada em *reporting* corporativos semelhantes, implementação de funcionalidades relevantes e ideias construtivas para uma melhor criação dos modelos de dados, o qual permite obter uma visão mais abrangente dos principais conceitos bem como das relações entre os mesmos;
- Habilitar utilizadores – Pretende-se o acompanhamento por parte da equipa de desenvolvimento e a participação do Cliente X. e seus interlocutores na manutenção futura das soluções e que, inclusive, venham a aplicar novamente as mesmas abordagens e ferramentas em futuras iniciativas de transformação;
- Gestão da mudança, em que os *dashboards* criados constituirão uma ferramenta de trabalho diária para os utilizadores atendendo ao âmbito de consulta de cada grupo de trabalho;
- Integrar novas fontes de informação e renovar o catálogo de visualizações e conteúdos (alargar o DW);
- Implementar redefinições do *Modelo de Governance*, ou seja, possibilidade de criação de relatórios e *dashboards* atendendo aos diferentes grupos de trabalho e respetivos âmbitos de análise bem como a integração de novos pedidos ou reformulação de pedidos antigos;
- Otimização dos modelos implementados, visando um maior detalhe de informação e novos modelos de negócio que, entretanto, possam surgir;
- Participação na manutenção evolutiva da plataforma de *reporting* corporativo do Cliente X.



# Referências Bibliográficas

As referências bibliográficas constituem uma listagem das referências citadas no corpo do projeto e estão por ordem alfanumérica pelo apelido do primeiro ou único autor, de acordo com as normas da referência APA (6ª edição).

Ariyachandara, T., Watson, H. (2010). *Key organizational factors in Datawarehouse architecture selection*. Decision Support Systems 49, 200–212.

Azevedo, A., Santos, M. F. (2009). *Business intelligence: State of the Art*. Trends, and Open Issues. In Proceedings of the First International Conference on Knowledge Management and Information Sharing - KMIS 2009 (pp.296-300).

BI LIVRE (2017). *A Evolução do Business Intelligence*. Disponível em: <http://www.bilivre.com.br/artigos/a-evolucao-do-business-intelligence/>.

Brunton, R. (2015). *Cloud, On Premise or Hybrid: Choosing the right solution for your business*. Disponível em: <https://www.maximizer.com/blog/Cloud-on-premise-or-hybrid-choosing-the-right-solution-for-your-business/>.

Castellanos, M., Dayal, U., Simitsis, A., Wilkinson, K. (2009). *Data Integration Flows for Business Intelligence*. In Proceedings of the International Conference on Extending Database Technology (EDBT), Saint Petersburg, Russia.

Chaudhuri, S., Dayal, U., Narasayya, V. (2011). *An Overview of Business Intelligence Technology*. Communications of the ACM, 88-98.

Collaborate in your Power BI app Workspace. (2017). Microsoft Docs. Disponível em: <https://docs.microsoft.com/en-us/power-bi/service-collaborate-power-bi-workspace>.

Dashboards no serviço Power BI. (2017). Microsoft Docs. Disponível em: <https://docs.microsoft.com/pt-pt/power-bi/service-dashboards>.

Descrição Geral da Linguagem DAX (*Data Analysis Expressions*). (2018). Microsoft Docs – TechNet. Disponível em: [https://technet.microsoft.com/pt-pt/library/gg399181\(v=sql.110\).aspx](https://technet.microsoft.com/pt-pt/library/gg399181(v=sql.110).aspx).

DirectQuery. (2018). Microsoft Docs. Disponível em: <https://docs.microsoft.com/pt-pt/power-bi/desktop-use-directquery#limitations-of-directquery>.

Gallo, J. (2018). *Self-Service BI: Barriers, Benefits and Best Practices*. TDWI. Disponível em: <https://tdwi.org/articles/2018/04/24/bi-all-self-service-bi-barriers-benefits-and-best-practices.aspx>.

Gardner, S. R. (1998). *Building the Datawarehouse*. Communications of the ACM, 52-60.

Golfarelli, M., Maio, D., Rizzi, S. (1998). *Conceptual Design of Datawarehouses from E/R Schemes*. Published in the Proceedings of the Hawaii International Conference On System Sciences, IEEE.

Haiyan, H., Zhenyuan, W. (2010). *OLAP Technology and Its Business Application*. 2010 Second WRI Global Congress on Intelligent Systems, 92-95.

Inmon, W. H. (2005). *Building the Datawarehouse*. Wiley, 4nd edition.

Juliano. (2008). *Entendendo a Linguagem SQL*. DEVMEDIA. Disponível em: <https://www.devmedia.com.br/entendendo-a-linguagem-sql/7775>.

Keunecke, V. (2016, dezembro 27). *Projetos Eficientes – Gráfico de Gantt na Gestão de Projetos como exemplo Prático*. Hinc. Disponível em: <https://hinc.com.br/grafico-de-gantt-gestao-projetos-exemplo>.

Kimball, R., Ross, M. (2013). *The Datawarehouse Toolkit: The Definitive Guide to Dimensional Modeling*, Terceira edição. Indianapolis, IN 46256. Publicado por John Wiley & Sons, Inc.

Master Data Services. (2014). Microsoft Docs. Disponível em: <https://docs.microsoft.com/en-us/sql/master-data-services/master-data-services?view=sql-server-2014>.

MicroStrategy University (2008). *MicroStrategy desktop: reporting essentials*. MicroStrategy Education course guide, Virginia, USA.

Nemec, J., Nguyen, T. M., Tjoa, A. M., Windisch, M. (2006). *An approach towards an event-fed solution for slowly changing dimensions in Datawarehouses with a detailed case study*. Data & Knowledge Engineering 63 (2007) Elsevier, 26-43.

Nussbaumer, H., Revels, M. (2013). Data Mining and Data Warehousing in the Airline Industry. *Academy of Business Research Journal*; 2013, Vol. 3, p69.

O que é o Power BI? (2017). Microsoft Docs. Disponível em: <https://docs.microsoft.com/pt-pt/power-bi/power-bi-overview>.

O que é o Power BI Desktop? (2017). Microsoft Docs. Disponível em: <https://docs.microsoft.com/pt-pt/power-bi/desktop-what-is-desktop>.

Pant, P. (2009). *Business intelligence (BI): How to build successful BI strategy*. Deloitte Consulting LLP.

Portugal, Constituição da República Portuguesa (2013). Diário da República. Artigo 5.º, Lei n.º 29/2013 de 19-04-2013 – Princípio da Legalidade. Portugal [s.n.] 2013.

Portugal, Constituição da República Portuguesa (2017). Diário da República. Portugal [s.n.] 2017. pp. 5232 (2.ª Série, N.º58).

Ramos, I., Santos, M. Y. (2009). *Business Intelligence – Tecnologias da Informação na Gestão de Conhecimento* (2 ed.). Lisboa, Portugal: FCA.

Rainardi, V. (2008). *Building a Datawarehouse: With Examples in SQL Server*. United States of America: Apress.

Ranjan, J. (2009). Business Intelligence: Concepts, Components, Techniques and Benefits. *Journal of Theoretical and Applied Information Technology*, 9(1). 60-70.

Relatórios no Power BI. (2017). Microsoft Docs. Disponível em: <https://docs.microsoft.com/pt-pt/power-bi/service-reports>.

Scrum Methodology – Learn Scrum. (2001). Disponível em: <http://scrummethodology.com/>.

Shaker H., Ali El-Sappagh, Abdeltawab, M., Ahmed, Hendawi, Ali Hamed El Bastawissy (2011). *A proposed model for Datawarehouse ETL processes*. Journal of King Saud University – Computer and Information Sciences (2011) 23, 91–104.

Silva, R (2014). Citações e Referências bibliográficas – Estilo APA (*American Psychological Society*). 6ª edição. Biblioteca ESTGV.

Simitsis, A., Skiadopoulos, S., Vassiliadis, P. (2002). *Conceptual Modeling for ETL Processes*. Paper presented at the 5th ACM international workshop on Data Warehousing and OLAP. Virgínia, USA.

SQL Server Management Studio (SSMS). (2017). Microsoft Docs. Disponível em: <https://docs.microsoft.com/en-us/sql/ssms/sql-server-management-studio-ssms?view=sql-server-2017>.

SQL Server Integration Services. (2017). Microsoft Docs. Disponível em: <https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-2017>.

Unipartner IT Services, S.A. (2018). *Unipartner IT Services – Your Partner in a Digital Future*. Disponível em: <https://www.unipartner.com/>.

Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. Politecnico di Milano, Italy: A John Wiley and Sons, Ltd., Publication.

Watson, H. J., Houdeshel G., Rainer, R. K. (1997). *Building Executive Information Systems and Other Decision Support Applications*, Wiley.



# Apêndices

## Apêndice I – Diagrama de Gantt

Estima-se o seguinte plano de trabalho e calendário de realização das atividades que inclui, de uma maneira geral, estabelecimento das bases que assegurem a execução do projeto com qualidade: diagnóstico e levantamento de requisitos, modelação, processos de ETL, produção e disponibilização de relatórios em Microsoft *Power BI*, de acordo com o plano geral do projeto supracitado, com a indicação das principais fases consideradas.

A realização do projeto teve conclusão em 33 semanas. De uma maneira geral, salientam-se as principais etapas: lançamento do projeto, análise funcional e desenho técnico, preparação da infraestrutura, processos de ETL, modelo analítico de dados das três fontes de dados referentes a atividade planeada e corrente, criação do portal de *reporting* corporativo e resposta aos relatórios mensais do Cliente X., testes de qualidade e de aceitação e arranque em produção. A implementação do projeto engloba diferentes atividades descritas em maior detalhe nas páginas que se seguem.

### • Descrição sumária do Diagrama de Gantt:

De uma maneira geral, o diagrama de Gantt engloba cinco principais fases: pré-requisitos da autora como estagiária; iniciação ao projeto 1, com uma entidade pública; desenvolvimento do projeto 1; mudança de projeto e desenvolvimento do projeto 2; operações que dão o projeto por concluído e, por fim, a última fase que diz respeito ao lançamento final da solução. De uma maneira geral, o diagrama de Gantt envolveu as seguintes atividades:

- *Start-up* – apresentação da empresa, da equipa, do projeto e objetivos;
- Reunião de *kick-off* com o Cliente X. – definição de objetivos e equipa de projeto;
- Estabelecimento dos procedimentos do projeto e condições logísticas;
- Calendarização das sessões de trabalho;
- Levantamento dos requisitos funcionais da solução a implementar (análise e desenho das funcionalidades);
- *Workshops* de levantamento de requisitos;
- Desenho funcional dos processos e operações a disponibilizar;
- Validação das regras de negócio a implementar;
- Definição da infraestrutura de suporte à solução;
- Desenho técnico da solução base e serviços transversais;
- Desenho técnico dos módulos e funcionalidades a disponibilizar;
- Desenho das interfaces de ligação aos sistemas do Cliente X.;
- Definição do dicionário de dados, repositórios de informação, e regras de segurança;
- Apoio na instalação e configuração do serviço *Power BI*;
- Apoio na identificação de requisitos/regras de rede, comunicações e segurança;

- Configuração da infraestrutura da solução, dos repositórios de informação e regras dos processos;
- Desenvolvimento dos processos de ETL e modelos analíticos – inclui, ainda, implementação de novas dimensões sobre os modelos de atividade;
- Desenvolvimento, implementação e configuração do Portal de *Reporting*;
- Configuração de relatórios e *dashboards*;
- Produção de 3 relatórios/*dashboards* por modelo analítico;
- Disponibilização de documentação técnica de suporte;
- Desenvolvimento/configuração dos mecanismos de exportação de dados dos sistemas operacionais – Fontes de dados: A, B e C;
- Realização de testes integrados e de desempenho da solução;
- Apoio remoto aos testes de aceitação e arranque da solução em produção, de acordo com a metodologia acordada;
- Produção de relatórios *self-service* através da plataforma *Power BI*.

• **Fase 1: Pré-Requisitos:**

Tabela 15: Fase 1 do Diagrama de Gantt – Pré-Requisitos

<u>Módulo</u>	<u>Atividade</u>	<u>Data de Início</u>	<u>Data de Fim</u>	<u>Duração (em dias)</u>
<b><u>Fase 1: Pré-requisitos</u></b>	<b>Atividade 01: <i>Start-up</i>: Acolhimento, apresentação da empresa e conhecimento do funcionamento da empresa; Aspetos Logísticos e ambientação à realidade da empresa.</b>	25/09/2017	25/09/2017	0d
	<b>Atividade 02: Integração na fase final de um projeto de BI:</b>	26/09/2017	17/11/2017	52d
	Atividade 2.1.: <i>Login</i> para as máquinas e explicação das funções exigidas para o projeto.	26/09/2017	26/09/2017	0d
	Atividade 2.2.: Exploração do <i>Visual Studio</i> em projetos de <i>Integration Services</i> e <i>Analysis Services</i> ; Criação de manuais de utilização em <i>Microsoft Power BI</i> ; Consulta de alguns vídeos da <i>Microsoft Academy</i> sobre <i>Data Analytics</i> para consolidação de conhecimentos.	26/09/2017	13/10/2017	17d
	Atividade 2.3.: Pesquisa de <i>Exchanged Rates</i> (AOA – Kwanza Angolano, BRL – Real Brasileiro e USD – Dólar Americano, to_EUR e EUR_to) de 2007 a 2017.	15/10/2017	20/10/2017	5d
	Atividade 2.4.: Construção do guião de apoio sobre a pesquisa de <i>Exchanged Rates</i> .	23/10/2017	23/10/2017	0d
	Atividade 2.5.: Construção de <i>pivot tables</i> sobre <i>Exchanged Rates</i> .	24/10/2017	27/10/2017	3d
	Atividade 2.6.: Integração de <i>Exchanged Rates</i> no <i>Microsoft Visual Studio</i> .	30/10/2017	01/11/2017	2d



Atividade 2.7.: Criação de um documento em Microsoft Excel de MDX ( <i>MultiDimensional eXpressions</i> ).	02/11/2017	10/11/2017	8d
Atividade 2.8.: Construção de um guião em Microsoft Power BI do percurso Cubo ->EDW ->Stating ->ODS -> SO.	03/11/2017	17/11/2017	14d
Atividade 2.9.: Término da contribuição da autora na fase final de um projeto de BI.	17/11/2017	17/11/2017	0d
<b>Atividade 03: Sessão de formação da Microsoft Partner University (conhecimento dos cursos e convívio entre os estagiários).</b>	17/11/2017	17/11/2017	0d
<b>Atividade 04: Reunião com o orientador da empresa (definição de resumo e plano de trabalho para entrega de registo de tese).</b>	29/11/2017	29/11/2017	0d
<b>Atividade 05: Reunião com o orientador da faculdade (definição de resumo e plano de trabalho para entrega de registo de tese).</b>	30/11/2017	30/11/2017	0d
<b>Atividade 06: Aquisição de conhecimento específico, através de <i>self-study</i> em plataformas de <i>e-learning/virtual academies</i> – Microsoft Partner University:</b>	20/11/2017	06/01/2018	47d
Atividade 6.1.: Office 2013 for IT Professionals (MPN10338).	20/11/2017	24/11/2017	4d
Atividade 6.2.: Virtual Machines in Azure (level 100) (MPN10726).	20/11/2017	24/11/2017	4d
Atividade 6.3.: Microsoft Azure course: .NET Development with Azure 101 (MPN18112).	20/11/2017	24/11/2017	4d
Atividade 6.4.: Introduction to Microsoft Dynamics CRM 2016.	20/11/2017	06/12/2017	16d
Atividade 6.5.: Devices and Deployment Technical for Deploying Desktop Virtualization Technologies.	20/11/2017	06/01/2018	47d
Atividade 6.6.: C+E University: IoT Online (MPN17160).	23/11/2017	23/11/2017	0d
Atividade 6.7.: Accelerate Package – Azure Fast Start for IaaS (Infrastructure as a Service) Foundation (MPN15797).	23/11/2017	24/11/2017	1d
Atividade 6.8.: Accelerate Package – Azure Fast Start for Website on PaaS (Platform as a Service) (MPN15856).	23/11/2017	24/11/2017	1d
Atividade 6.9.: Microsoft Dynamics 365 – Foundation.	23/11/2017	24/11/2017	1d

Atividade 6.10.: <i>Accelerate – Microsoft Azure Fast Start for Modern Application Development</i> (MPN16023).	24/11/2017	24/11/2017	0d
Atividade 6.11.: <i>Accelerate Package – Azure Fast Start for Website on IaaS</i> (MPN15855).	24/11/2017	27/11/2017	3d
Atividade 6.12.: <i>C+E University: Cloud Application Development Online</i> (MPN16885).	24/11/2017	27/11/2017	3d
Atividade 6.13.: <i>Microsoft Dynamics 365 – Intelligent Business Applications</i> (MPN17001).	24/11/2017	27/11/2017	3d
Atividade 6.14.: <i>Accelerate Package – Azure Fast Start for SQL on IaaS</i> (MPN16855).	24/11/2017	29/11/2017	5d
Atividade 6.15.: <i>Accelerate – Azure Fast Start for Business Analytics with Power BI</i> (MPN17058).	24/11/2017	29/11/2017	5d
Atividade 6.16.: <i>C+E University: Enterprise Mobility Suite</i> (MPN16622).	24/11/2017	01/12/2017	7d
Atividade 6.17.: <i>C+E University: Cloud Infrastructure</i> (MPN16867).	24/11/2017	04/12/2017	10d

Fonte: Elaborado pela autora

- **Fase 2: Iniciação:**

Tabela 16: Fase 2 do Diagrama de Gantt – Iniciação

<b><u>Módulo</u></b>	<b><u>Atividade</u></b>	<b><u>Data de Início</u></b>	<b><u>Data de Fim</u></b>	<b><u>Duração (em dias)</u></b>
<b><u>Fase 2: Iniciação</u></b>	<b>Atividade 01: Aspectos logísticos e ambientação ao Cliente Y.</b>	06/11/2017	06/11/2017	0d
	<b>Atividade 02: Apresentação da autora à equipa.</b>	06/11/2017	06/11/2017	0d
	<b>Atividade 03: Apresentação do Projeto: Plataforma de Gestão de Processos e de <i>Business Intelligence</i>.</b>	07/11/2017	07/11/2017	0d
	<b>Atividade 04: Enquadramento global do projeto (introdução ao projeto e seus objetivos).</b>	07/11/2017	10/11/2017	3d
	<b>Atividade 05: Análise da documentação existente, do problema e estruturação da metodologia a utilizar.</b>	07/11/2017	10/11/2017	3d
	<b>Atividade 06: Reunião com a equipa de trabalho.</b>	10/11/2017	10/11/2017	0d
	<b>Atividade 07: Início do projeto.</b>	14/11/2017	14/11/2017	0d

	<b>Atividade 08: Entrega da documentação para o registro da tese (resumo, plano de trabalho e <i>Curriculum Vitae</i> do orientador da empresa) no Gabinete de Estudos Pós-Graduados (GEPG).</b>	13/12/2017	13/12/2017	0d
	<b>Atividade 09 (contínua): Redação e entrega do relatório preliminar sobre as atividades desenvolvidas.</b>	14/11/2017	05/01/2018	52d

Fonte: Elaborado pela autora

- **Fase 3: Desenvolvimento Parte I:**

Tabela 17: Fase 3 do Diagrama de Gantt – Desenvolvimento Parte I

<u>Módulo</u>	<u>Atividade</u>	<u>Data de Início</u>	<u>Data de Fim</u>	<u>Duração (em dias)</u>
<b><u>Fase 3:</u></b> <b><u>Desenvolvimento</u></b>	<b>Atividade 01: Elaboração do Projeto 1:</b>	14/11/2017	31/07/2018	259d
	Atividade 1.1.: Diagnóstico: conhecimento dos processos de negócio, aplicações e infraestruturas tecnológicas.	01/08/2016	01/08/2017	365d
	Atividade 1.2.: Levantamento dos processos específicos, comuns e transversais.	01/08/2017	01/09/2017	31d
	Atividade 1.3.: Definição do modelo operativo de referência, avaliação da situação atual, caracterização das diferenças ( <i>gaps</i> ) e prioridades de ação ( <i>roadmap</i> ).	01/09/2017	01/12/2017	91d
	Atividade 1.4.: Modelação de processos com BPMN ( <i>Business Process Model and Notation</i> ) 2.0.	01/09/2017	01/12/2017	91d
	Atividade 1.5.: Levantamento da arquitetura atual.	01/09/2017	01/12/2017	91d
	Atividade 1.6.: Definição da arquitetura empresarial de negócio (atual e de referência) em ArchiMate 3.0.	01/09/2017	01/12/2017	91d
	Atividade 1.7.: Definição da arquitetura empresarial aplicacional (atual e de referência) em ArchiMate 3.0.	01/09/2017	01/12/2017	91d
	Atividade 1.8.: Definição da arquitetura empresarial tecnológica (atual e de referência) em ArchiMate 3.0.	01/09/2017	01/12/2017	91d
	Atividade 1.9.: Identificação e caracterização dos indicadores de desempenho e risco, metas e métricas dos processos.	01/09/2017	31/12/2017	121d
	Atividade 1.10.: Estudo e implementação dos processos de negócio futuros.	01/09/2017	01/12/2017	91d
	Atividade 1.11.: Reuniões pontuais de validação da modelação e arquitetura empresarial dos projetos.	01/09/2017	13/11/2017	73d
	Atividade 1.12.: Sessões para definir especificações pontuais dos formulários em Fabasoft eGov-Suite.	14/11/2017	09/01/2018	56d
	Atividade 1.13.: Parametrização da plataforma Fabasoft eGov-Suite para suportar requisitos do Cliente F.	19/12/2017	31/01/2018	43d
	Atividade 1.14.: Desmaterialização dos processos comuns e transversais.	09/01/2018	06/03/2018	56d
	Atividade 1.15.: Implementação dos formulários na plataforma Fabasoft eGov-Suite.	09/01/2018	06/03/2018	56d
	Atividade 1.16.: Revisão do catálogo de serviços, segundo os critérios definidos para os processos desmaterializados.	06/03/2018	06/03/2018	0d
	Atividade 1.17.: Reunião para definir plano para projetos 3 e 4.	30/04/2018	30/04/2018	0d

	<b>Mudança de Projeto.</b>	12/12/2017	31/05/2018	170d
--	----------------------------	------------	------------	------

Fonte: Elaborado pela autora

- **Fase 3: Desenvolvimento Parte II:**

Nesta fase, ocorre a mudança do projeto 1 referente a uma entidade pública para o projeto 2 referente a uma entidade bancária.

Tabela 18: Fase 3 do Diagrama de Gantt – Desenvolvimento Parte II

<u>Módulo</u>	<u>Atividade</u>	<u>Data de Início</u>	<u>Data de Fim</u>	<u>Duração (em dias)</u>
<b>Fase 3:</b> <b><u>Desenvolvimento</u></b>	<b>Atividade 01: Preparação do Projeto 2:</b>	02/01/2018	02/01/2018	0d
	Atividade 1.1.: Apresentação do Projeto: Portal de <i>Reporting</i> (Power BI).	02/01/2018	05/01/2018	3d
	Atividade 1.2.: Enquadramento global do projeto (introdução ao projeto, seus objetivos, análise do problema e estruturação da metodologia a utilizar).	02/01/2018	05/01/2018	3d
	Atividade 1.3.: Análise da documentação existente do projeto.	02/01/2018	05/01/2018	3d
	Atividade 1.4.: Reunião com o gestor de projeto e a equipa de trabalho.	03/01/2018	03/01/2018	0d
	Atividade 1.5.: Reunião com o orientador da faculdade para rever conceitos do relatório preliminar; Definição de fios condutores para o projeto; <i>Briefing</i> sobre novo projeto da empresa X.; Comunicação de mudança de projeto e plano ao Professor Luís Moniz, responsável pelas teses de mestrado FCUL.	04/01/2018	04/01/2018	0d
	Atividade 1.6.: Entrega do relatório preliminar referente ao projeto anterior; Análise da proposta do novo projeto; Esquematização das tarefas. Análise da documentação necessária referente à fase inicial do projeto.	05/01/2018	05/01/2018	0d
	<b>Atividade 02: Início do projeto:</b>	12/12/2017	05/01/2018	24d
	Atividade 2.1.: <i>Kick-Off</i> do Projeto BI - Informação de Gestão.	12/12/2017	12/12/2017	0d
	Atividade 2.2.: Análise Funcional <i>High-Level</i> (Tratamento e Modelação da informação existente).	12/12/2017	19/01/2018	38d
	<b>1ª Release – Desenvolvimento.</b>	18/12/2017	29/01/2018	42d
	<b>Atividade 03: Construção da 1ª Release do Relatório Mensal:</b>	18/12/2017	24/01/2018	37d
	Atividade 3.1.: Conhecimento das instalações e funcionamento da dinâmica do Cliente X. pela autora; acessos a máquinas e computador de trabalho.	08/01/2018	08/01/2018	0d
	Atividade 3.2.: Diagnóstico: Levantamento de requisitos das três fontes operacionais.	08/01/2018	12/01/2018	4d

Atividade 3.3.: Modelo <i>SandBox</i> da Fonte de dados A + Relatórios.	08/01/2018	19/01/2018	11d
Atividade 3.4.: Modelo <i>SandBox</i> da Fonte de dados B + Relatórios.	08/01/2018	19/01/2018	11d
Atividade 3.5.: Modelo <i>SandBox</i> da Fonte de dados C + Relatórios.	08/01/2018	19/01/2018	11d
Atividade 3.6.: Sessão de esclarecimento dos campos/colunas que pertencem ao documento em excel da fonte de dados A.	08/01/2018	08/01/2018	0d
Atividade 3.7.: Ponto de Situação Geral do Projeto.	08/01/2018	08/01/2018	0d
Atividade 3.8.: Implementação do modelo inicial para a máquina virtual.	11/01/2018	11/01/2018	0d
Atividade 3.9.: Ponto de Situação do Projeto (plano de trabalho + atividades a desenvolver).	15/01/2018	15/01/2018	0d
Atividade 3.10.: Ponto de Situação Geral do Projeto.	17/01/2018	17/01/2018	0d
Atividade extra: Assinatura do Acordo Específico a formalizar o trabalho de projeto.	17/01/2018	17/01/2018	0d
Atividade 3.11.: Ponto de Situação do Projeto (definição da Arquitetura da Solução).	18/01/2018	18/01/2018	0d
Atividade 3.12.: Validação da abordagem implementada e das ferramentas.	19/01/2018	19/01/2018	0d
Atividade 3.13.: Conclusão da execução dos relatórios e entrega do produto final fase 1.	19/01/2018	19/01/2018	0d
<b>Atividade 04: Análise Funcional Detalhada:</b>	22/01/2018	29/01/2018	7d
Atividade 4.1.: Ponto de Situação Interno do Projeto (Preparação dos <i>Workshops</i> ).	22/01/2018	22/01/2018	0d
Atividade 4.2.: <i>Workshop</i> revisão requisitos da fonte de dados A + Segurança.	23/01/2018	23/01/2018	0d
Atividade 4.3.: <i>Workshop</i> revisão requisitos da fonte de dados C + Segurança.	25/01/2018	25/01/2018	0d
Atividade 4.4.: <i>Workshop</i> revisão requisitos da fonte de dados B + Segurança.	29/01/2018	29/01/2018	0d
Atividade 4.5.: Revisão do Documento Técnico (definição dos mecanismos, contas de acesso, nomenclaturas e tabelas de mapeamento).	29/01/2018	29/01/2018	0d
Atividade 4.6.: Recolher e modelar informação adicional de negócio, resultante do <i>feedback</i> dos pontos de situação.	29/01/2018	29/01/2018	0d
Atividade 4.7.: Escrita do documento funcional da fonte de dados A.	30/01/2018	16/02/2018	17d
Atividade 4.8.: Escrita do documento funcional da fonte de dados B.	30/01/2018	16/02/2018	17d
Atividade 4.9.: Escrita do documento funcional da fonte de dados C.	30/01/2018	16/02/2018	17d
Atividade 4.10.: Validação dos modelos atuais por comparação com os antigos.	30/01/2018	30/01/2018	0d
Atividade 4.11.: Consolidação de Informação.	31/01/2018	31/01/2018	0d
<b>2ª Release – Desenvolvimento.</b>	01/02/2018	04/05/2018	92d
Análise de incongruências e inconsistências de dados, correção de dados, decisão da	01/02/2018	05/02/2018	4d

	entrada de novos campos para as várias fontes de dados e preparação da fase 2.			
	<b>Atividade 05: Definição da Arquitetura:</b>	05/02/2018	06/04/2018	60d
	Atividade 5.1.: Definição da Arquitetura de Desenvolvimento, Qualidade e Produção.	15/01/2018	15/01/2018	0d
	Atividade 5.2.: Aprovisionamento do Ambiente de Desenvolvimento.	05/02/2018	06/04/2018	60d
	Atividade 5.3.: Aprovisionamento do Ambiente de Qualidade.	16/02/2018	06/04/2018	49d
	Atividade 5.4.: Aprovisionamento do Ambiente de Produção.	23/02/2018	06/04/2018	42d
	Atividade 5.5.: Revisão da análise funcional.	23/01/2018	06/04/2018	73d
	Atividade 5.6.: Validação da arquitetura da solução a ser implementada.	12/02/2018	12/02/2018	0d
	<b>Atividade 06: Camada de Ingestão de Dados:</b>	19/01/2018	20/04/2018	91d
	Atividade 6.1.: Ponto de Situação do Projeto (definição dos campos de análise das várias fontes).	05/02/2018	02/04/2018	56d
	Atividade 6.2.: Construção da <i>Staging, Datawarehouse</i> .	23/02/2018	02/04/2018	38d
	Atividade 6.3.: Desenvolvimento da camada ETL e repositório da fonte de dados A.	30/01/2018	09/02/2018	10d
	Atividade 6.4.: Desenvolvimento da camada ETL e repositório da fonte de dados B.	01/02/2018	01/02/2018	0d
	Atividade 6.5.: Desenvolvimento da camada ETL e repositório da fonte de dados C.	30/01/2018	09/02/2018	10d
	Atividade 6.6.: Ponto de Situação Geral do Projeto.	30/01/2018	09/02/2018	10d
	Atividade 6.7.: Integração do histórico do Fonte de dados A com as regras recebidas pelo Cliente X.	30/01/2018	09/02/2018	10d
	<b>Atividade 07: Camada de Processamento de Dados:</b>	30/01/2018	09/02/2018	10d
	Atividade 7.1.: Ponto de Situação Geral do Projeto.	08/02/2018	08/02/2018	0d
	Atividade 7.2.: Ponto de Situação do Projeto (definição do modelo de dados).	12/02/2018	16/02/2018	4d
	<b>Atividade 08: Camada de Repositório de Dados:</b>	12/02/2018	16/02/2018	4d
	Atividade 8.1.: Ponto de Situação Geral do Projeto.	15/02/2018	15/02/2018	0d
	Atividade 8.2.: Modelo Analítico Atividade Planeada – Fonte de dados A + Fonte de dados B.	16/02/2018	16/02/2018	0d
	Atividade 8.3.: Modelo Analítico Atividade Corrente – Fonte de dados C.	19/02/2018	02/03/2018	11d
	Atividade 8.4.: Conclusão das componentes de Transformação e Carregamento.	22/02/2018	22/02/2018	0d
	<b>Atividade 09: Camada de Apresentação – Portal de Reporting (Power BI):</b>	19/04/2018	03/05/2018	14d
	Atividade 9.1.: Ponto de Situação do Projeto (definição das tabelas de mapeamento + campos de análise das tabelas).	19/04/2018	03/05/2018	14d

	Atividade 9.2.: <i>Power BI Governance</i> .	19/04/2018	03/05/2018	14d
	Atividade 9.3.: Relatórios finais – fonte de dados C.	05/03/2018	16/03/2018	11d
	Atividade 9.4.: Relatórios finais – Fonte de dados A.	08/03/2018	08/03/2018	0d
	Atividade 9.5.: Relatórios finais – Fonte de dados B.	05/03/2018	16/03/2018	11d
	Atividade 9.6.: <i>Dashboards</i> Consolidados pelo departamento de planeamento e controlo do Cliente X.	05/03/2018	16/03/2018	11d
	Atividade 9.7.: Ponto de Situação Interno do Projeto (Passagem de conhecimento sobre Relatórios em <i>Power BI</i> ).	05/03/2018	16/03/2018	11d
	Atividade 9.8.: Ponto de Situação do Projeto (primeira entrega dos Relatórios finais em <i>Power BI</i> ).	05/03/2018	16/03/2018	11d
	Atividade 9.9.: Ponto de Situação do Projeto (continuação da primeira entrega dos Relatórios finais em <i>Power BI</i> ).	05/03/2018	16/03/2018	11d
	Atividade 9.10.: Ponto de Situação Geral do Projeto.	19/03/2018	19/03/2018	0d
	Atividade 9.11.: Ponto de Situação Geral do Projeto (continuação).	29/03/2018	29/03/2018	0d
	Atividade 9.12.: Entrega dos Relatórios finais da fonte de dados B.	04/04/2018	04/04/2018	0d
	Atividade 9.13.: Análise da fonte de dados A Finalizada/ Aprovada.	05/04/2018	05/04/2018	0d
	Atividade 9.14.: Análise da fonte de dados B Finalizada/ Aprovada.	09/04/2018	09/04/2018	0d
	Atividade 9.15.: Análise da fonte de dados C Finalizada/ Aprovada.	12/04/2018	12/04/2018	0d
	Atividade 9.16.: Definição das contas de acesso da fonte de dados A (1 ficheiro .csv).	12/04/2018	12/04/2018	0d
	Atividade 9.17.: Definição das contas de acesso + definição das contas de acesso e local ao excel da fonte de dados B (ficheiro excel + base de dados).	12/04/2018	12/04/2018	0d
	Atividade 9.18.: Definição das contas de acesso da fonte de dados C (base de dados).	12/04/2018	12/04/2018	0d
	Atividade 9.19.: Definição das contas de acesso MDS (tabelas de mapeamento).	12/04/2018	12/04/2018	0d
	<b>Atividade 10: Instalação e Disponibilização em Qualidade:</b>	23/04/2018	04/05/2018	11d
	Atividade 10.1.: Desenvolvimento dos processos automáticos de extração de dados.	23/04/2018	04/05/2018	11d
	Atividade 10.2.: Procedimento de instalação incluindo <i>scripts</i> de instalação.	23/04/2018	04/05/2018	11d
	Atividade 10.3.: Eventuais Ajustes (incorporação do <i>feedback</i> recebido pelo requisitante do projeto).	23/04/2018	04/05/2018	11d
	<b>Atividade 11: Instalação e Disponibilização em Produção:</b>	07/05/2018	07/05/2018	0d
	Atividade 11.1.: Entrada em Produção.	07/05/2018	07/05/2018	0d
	Atividade 11.2.: Apoio Pós-produção.	08/05/2018	31/05/2018	23d

Fonte: Elaborado pela autora

- **Fase 4: Operações:**

Tabela 19: Fase 4 do Diagrama de Gantt – Operações

<u>Módulo</u>	<u>Atividade</u>	<u>Data de Início</u>	<u>Data de Fim</u>	<u>Duração (em dias)</u>
<b><u>Fase 4: Operações</u></b>	<b>Atividade 01: Execução de atividades de instalação, configuração de ambientes. Testes de soluções e correções do sistema implementado.</b>	07/05/2018	25/05/2018	18d
	<b>Atividade 02: Configurações e conclusões.</b>	20/05/2018	31/05/2018	11d

Fonte: Elaborado pela autora

- **Fase 5: Lançamento:**

Tabela 20: Fase 5 do Diagrama de Gantt – Lançamento

<u>Módulo</u>	<u>Atividade</u>	<u>Data de Início</u>	<u>Data de Fim</u>	<u>Duração (em dias)</u>
<b><u>Fase 5: Lançamento</u></b>	<b>Atividade 01: Redação do relatório final e respetivo resumo executivo para debate e apresentação.</b>	05/01/2018	24/09/2018	262d
	<b>Atividade 02: Entrega do relatório final.</b>	26/09/2018	26/09/2018	0d
	<b>Atividade 03: Entrega da documentação para admissão a provas públicas.</b>	26/09/2018	26/09/2018	0d
	<b>Atividade 04: Apresentação à equipa de projeto do realizado ao longo do projeto, lições aprendidas, aspetos positivos e menos positivos.</b>	01/10/2018	01/10/2018	0d
	<b>Atividade 05: Apresentação do relatório final, perante um júri, na faculdade.</b>	Data por definir	Data por definir	——

Fonte: Elaborado pela autora



## Apêndice II – Outras Dimensões existentes na Fonte de Dados A

- Dimensão “Stream”:

Tabela 21: Campos da Dimensão “Stream”

Campo	Tipo	Chave	SCD	Descrição do Atributo	Exemplo
PKStream	<i>Integer</i>	PK	-	Chave primária (invisível)	“1”
Stream	<i>Varchar (500)</i>	-	1	Stream (visível)	“PMO 2016 – Stream 3”

Fonte: Elaborado pela autora

- Dimensão “Contabilização”:

Tabela 22: Campos da Dimensão “Contabilização”

Campo	Tipo	Chave	SCD	Descrição do Atributo	Exemplo
PKAccounting	<i>Integer</i>	PK	-	Chave primária (invisível)	“1”
Accounting	<i>Varchar (500)</i>	-	1	Contabilização (visível)	“Gasto”

Fonte: Elaborado pela autora

- Dimensão “Plano”:

Tabela 23: Campos da Dimensão “Plano”

Campo	Tipo	Chave	SCD	Descrição do Atributo	Exemplo
PKPlan	<i>Integer</i>	PK	-	Chave primária (invisível)	“4”
Plan	<i>Varchar (500)</i>	-	1	Plano (visível)	“PDOT”

Fonte: Elaborado pela autora

- Dimensão “Programa”:

Tabela 24: Campos da Dimensão “Programa”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKProgram	<i>Integer</i>	PK	-	Chave primária (invisível)	“1”
Program	<i>Varchar (500)</i>	-	1	Programa (visível)	“Normalização de Processos”

Fonte: Elaborado pela autora

- Dimensão “Gestor de Projecto”:

Tabela 25: Campos da Dimensão “Gestor de Projecto”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKProjectManager	<i>Integer</i>	PK	-	Chave primária (invisível)	“2”
ProjectManager	<i>Varchar (500)</i>	-	1	Gestor de Projecto (visível)	“Andreia Dias”

Fonte: Elaborado pela autora

- Dimensão “Número de Projecto”:

Tabela 26: Campos da Dimensão “Número de Projecto”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKProjectNumber	<i>Integer</i>	PK	-	Chave primária (invisível)	“2”
ProjectNumber	<i>Varchar (500)</i>	-	1	Número de Projecto (visível)	“Factura-Recibo Anual”

Fonte: Elaborado pela autora

- Dimensão “Motivo”:

Tabela 27: Campos da Dimensão “Motivo”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKReason	<i>Integer</i>	PK	-	Chave primária (invisível)	“3”
Reason	<i>Varchar (500)</i>	-	1	Motivo (visível)	“Tecnológico”

Fonte: Elaborado pela autora

- Dimensão “Gestor da Relação”:

Tabela 28: Campos da Dimensão “Gestor da Relação”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKRelationshipManager	<i>Integer</i>	PK	-	Chave primária (invisível)	“1”
RelationshipManager	<i>Varchar (500)</i>	-	1	Gestor da Relação (visível)	“Miguel Vinagre”

Fonte: Elaborado pela autora

- Dimensão “Tipo de Pedido”:

Tabela 29: Campos da Dimensão “Tipo de Pedido”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKRequestType	<i>Integer</i>	PK	-	Chave primária (invisível)	“8”
RequestType	<i>Varchar (50)</i>	-	1	Tipo de Pedido (visível)	“Seleção de Software”

Fonte: Elaborado pela autora

- Dimensão “Dimensionamento”:

Tabela 30: Campos da Dimensão “Dimensionamento”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>SCD</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKSizingHour	<i>Integer</i>	PK	-	Chave primária (invisível)	“1”
SizingHour	<i>Varchar (500)</i>	-	1	Dimensionamento (visível)	“Pedido de Desenvolvimento Àgil”

Fonte: Elaborado pela autora

## Apêndice III – Outras Dimensões existentes na Fonte de Dados B

- Dimensão “Recurso”:

Tabela 31: Campos da Dimensão “Recurso”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKResource	<i>Integer</i>	PK	Chave Primária (invisível)	“1”
ResourceUID	<i>Varchar (255)</i>	-	Código do Recurso (visível)	“6FEF47B4-F247-E411-93F9-005056816BF8”
ResourceName	<i>Varchar (255)</i>	-	Nome do Recurso (visível)	“Rui Arrenega”
ResourceEmailAddress	<i>Varchar (255)</i>	-	Correio Eletrónico do Recurso (visível)	“x”
ResourceGroup	<i>Varchar (255)</i>	-	Grupo do Recurso (visível)	“GPO – Gab. de Gestão da Procura e da Oferta”
ResourceBaseCalendar	<i>Varchar (255)</i>	-	Calendário base do Recurso (visível)	“Externos”
TeamName	<i>Varchar (255)</i>	-	Nome da Equipa (visível)	“ARP”
ResourceCompany	<i>Varchar (255)</i>	-	Companhia do Recurso (visível)	“Unipartner”
ResourceStatus	<i>Varchar (255)</i>	-	Estado do Recurso (visível)	“In”
Organization	<i>Varchar (255)</i>	-	Organização (visível)	“x”
FKResourceType	<i>Integer</i>	FK	Este campo está invisível	“2”
FKRBS	<i>Integer</i>	FK	Este campo está invisível	“24”

Fonte: Elaborado pela autora

- Dimensão “Projecto”:

Tabela 32: Campos da Dimensão “Projecto”

<b>Campo</b>	<b>Tipo</b>	<b>Chave</b>	<b>Descrição do Atributo</b>	<b>Exemplo</b>
PKProject	<i>Integer</i>	PK	Chave Primária (invisível)	“1”

ProjectUID	<i>Varchar (255)</i>	-	Código de Projecto (visível)	“1F862500-71FE-E611-82CB-0024D691439C”
ProjectName	<i>Varchar (255)</i>	-	Nome do Projecto (visível)	“PR17009-Reformulação da Central de Responsabilidades de x”
ProjectDescription	<i>Varchar (512)</i>	-	Descrição do Projecto (visível)	“PR17009-Reformulação da Central de Responsabilidades de x”
ProjectStartDate	<i>Date</i>	-	Data de Início do Projecto (visível)	“2017-02-06”
ProjectFinishDate	<i>Date</i>	-	Data de Fim do Projecto (visível)	“2017-11-29”
ExpectedEnddate	<i>Date</i>	-	Data Esperada de Fim do Projecto (visível)	“2017-11-29”
ProjectTitle	<i>Varchar (255)</i>	-	Título do Projecto (visível)	“PR17009-Reformulação da Central de Responsabilidades de x”
ProjectCreatedDate	<i>Date</i>	-	Data de Criação do Projecto (visível)	“2017-03-01”
ProjectOwnerName	<i>Varchar (255)</i>	-	Nome do criador do Projecto (visível)	“Rui Arrenega”
FKClient	<i>Integer</i>	FK	-	“28”
FKServiceLevel	<i>Integer</i>	FK	-	“11”
ProjectStatus	<i>Varchar (255)</i>	-	Estado do Projecto (visível)	“Terminado”
ProjectManager	<i>Varchar (255)</i>	-	Gestor do Projecto (visível)	“Rui Arrenega”
ActivityPlan	<i>Varchar (255)</i>	-	Plano de Actividade do Projecto (visível)	“PPDOTR”
ProjectType	<i>Varchar (255)</i>	-	Tipo de Projecto (visível)	“Projecto”
Acronym	<i>Varchar (255)</i>	-	Sigla do Projecto (visível)	“PR”
Description	<i>Varchar (255)</i>	-	Descrição (visível)	“Projecto”
ActivityType	<i>Varchar (50)</i>	-	Tipo de Actividade (visível)	“Actividade Planeada”
RequestNumber	<i>Varchar (255)</i>	-	Número do Pedido associado ao Projecto (visível)	“2017/0036”

Fonte: Elaborado pela autora

## Apêndice IV – Campos para a Tabela de Mapeamento para o MDS

Tabela 33: Tabelas de Mapeamento para o MDS

Tabelas de Mapeamento			
Fonte de Dados	Tabela	Colunas	Tipo de Dados
Fonte de dados A	“Estados”	“Estados”	<i>Varchar (50)</i>
		“Sub_Estados”	<i>Varchar (50)</i>
		“Sub_Sub_Estados”	<i>Varchar (50)</i>
Fonte de dados B	“Projecto_Categorização”	“Tipo de Actividade”	<i>Varchar (50)</i>
		“Acrónimo”	<i>Varchar (255)</i>
		“Descrição”	<i>Varchar (50)</i>
Fonte de dados C	“ServiceType”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
	“Priority”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
	“Impact”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
	“Urgency”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
	“IncidentStatus”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
		“Estado Finito”	<i>Bit</i>
		“Ativo”	<i>Bit</i>
	“WorkordersStatus”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
		“Estado Finito”	<i>Bit</i>
		“Ativo”	<i>Bit</i>
	“Status_Reason_Incidents”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
	“Status_Reason_Workorders”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
	“X_ReportedSource/reported_Source”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
	“X_MotivoEstadoFornecedor”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
	“Work_Order_Type”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>
	“Vendor_Assignment_Status”	“Código”	<i>Integer</i>
		“Descrição”	<i>Varchar (50)</i>

Fonte: Elaborado pela autora